SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(85th Edition)

JANUARY 2020

BY: CENTRAL LIBRARY

DELHI TECHNOLOGICAL UNIVERSITY
(FORMERLY *DELHI COLLEGE OF ENGINEERING)*
GOVT. OF N.C.T. OF DELHI
SHAHBAD DAULATPUR, MAIN BAWANA ROAD
DELHI 110042

PREFACE

This is the **Eighty Fifth** Issue of Current Awareness Bulletin started by Delhi Technological University, Central Library. The aim of the bulletin is to compile, preserve and disseminate information published by the faculty, students and alumni for mutual benefits. The bulletin also aims to propagate the intellectual contribution of Delhi Technological University (DTU) as a whole to the academia.

The bulletin contains information resources available in the internet in the form of articles, reports, presentations published in international journals, websites, etc. by the faculty and students of DTU. The publications of faculty and student which are not covered in this bulletin may be because of the reason that the full text either was not accessible or could not be searched by the search engine used by the library for this purpose.

The learned faculty and students are requested to provide their uncovered publications to the library either through email or in CD, etc. to make the bulletin more comprehensive.

This issue contains the information published during **January, 2020**. The arrangement of the contents is alphabetical. The full text of the article which is either subscribed by the university or available in the web is provided in this bulletin.

Central Library

CONTENTS

1. A Study in Multistage Decision Making for Odisha State Avionic Infrastructure Development, ***3.L. Das*** and Chittaranjan Mallick, Mathematics, DTU

2. A survey on database intrusion detection: approaches, challenges and application, ***3.Rajni Jindal*** and ***3.Indu Singh***, Computer Science, DTU

3. Addition of Anti-Flaming Agents in Castor Oil based Rigid Polyurethane Foams: Studies on Mechanical and Flammable Behaviour, ***3.Raminder Kaur*** and ***6.Mukesh Kumar***, Chemistry, DTU

4. Analysis of mechanical behaviour of L-arginine Hydrobromide Monohydrate (LAHBr) single crystal grown by unidirectional growth technique, ***6.Kanika Thukral***, Sonia, Neema Jha, N. Vijayan, Budhendra Singh and S. A. Martin Britto Dhas, Applied Physics, DTU

5. Captivating Brand Hate Using Contemporary Metrics: A Structural Equation Modelling Approach, Richa Joshi and ***3.Rajan Yadav***, DSM, DTU

6. Comparative Study of Improvements in Multiple Effect Vapor Absorption Systems using Loop Heat Pipes, ***6.Ankit Dwivedi*** and ***3.R S Mishra***, Mechanical, DTU

7. Conducting Polymer Based Gas Sensor Using PNIN- Gate All Around - Tunnel FET, Jaya Madan, Rahul Pandey and ***3.Rishu Chaujar***, Applied Physics, DTU

8. Corporate entrepreneurship in India: The road less travelled, ***3.S. S. Khanka***, USME, DTU

9. DEEP LEARNING BASED ANDROID MALWARE DETECTION FRAMEWORK, Soumya Sourav, Devashish Khulbe and ***8.Naman Kapoor***, Computer Science, DTU

10. Design and Analysis of a Multi-band Flower Shaped Patch Antenna for WLAN/WiMAX/ISM Band Applications, ***6.Karteek Viswanadha*** and ***3.Nallanthighal Srinivasa Raghava***, Electronics, DTU

11. Parboiled Paddy Drying with Dierent Dryers: Thermodynamic and Quality Properties, Mathematical Modeling Using ANNs Assessment, Ebrahim Taghinezhad, Antoni Szumny, Mohammad Kaveh, Vali Rasooli Sharabiani, ***3.Anil Kumar*** and Naoto Shimizu, Mechanical, DTU

12. Domination in Signed Petri Net, ***3.Payal*** and ***3.Sangita Kansal***, A Mathematics, DTU

13. Effect of electrostatic precipitator on exhaust emissions in biodiesel fuelled CI engine, ***6.Ankit Sonthalia***, Shivam Garg, Rishav Sharma, Thiyagarajan Subramanian and ***3.Naveen Kumar***, Mechanical, DTU

14. Effect of Using Waste Plastic Oil on the Performance and Emissions Characteristics of Unmodified Constant Speed Direct Injection Diesel Engine, ***6.Parvesh Kumar***, P. Srinivasulu and ***6.Sidharth***, Mechanical, DTU

15. Electrical power system transmission quality and power supplier micro grid control functional reliability, ***3.L. N. Das*** and ***8.Sanyam Gupta***, Applied Mathematics, DTU

16. Empowering Women Through Financial Inclusion: A Study of Urban Slum, ***6.Shivangi Bhatia*** and ***3.Seema Singh***, Humanities, DTU

17. Ensemble-based extreme learning machine model for occupancy detection with ambient attributes, Sachin Kumar, ***3.Jagvinder Singh*** and Ompal Singh, Management, DTU

18. EPR and Optical Properties of Green Emitting Mn Activated Sr2ZnSi2O7 Phosphors Prepared by Sol–Gel Method, VIJAY SINGH, NISHA DEOPA, ***6.SUMANDEEP KAUR***, ***3.A.S. RAO***, J.L. RAO and G. LAKSHMINARAYANA, A Physics, DTU

19. Estimating maintenance budget using Monte Carlo simulation, ***3.Atul Kumar Srivastava***, ***3.Girish Kumar*** and Piyush Gupta, Civil, Mechanical, DTU

20. Exploring Barriers of Corporate Social Responsibility: Indian SMEs Perspective, Dr. Mohd Imran Khan, Shahbaz Khan and ***3.Mohd Shuaib***, Mechanical, DTU

21. Gratitude and leadership in higher education institutions Exploring the mediating role of social intelligence among teachers, 3.Naval Garg and ***6.Shilpika Gera***, USME, DTU

22. How Can Small and Medium Enterprises Effectively Implement Corporate Social Responsibility?: An Indian Perspective, ***6.Gaurav Dawar*** and ***3.Seema Singh***, Humanities, DTU

23. Investigation of physicochemical properties of oil palm biomass for evaluating potential of biofuels production via pyrolysis processes, Pranshu Shrivastava1, Phonthip Khongphakdi, Arkom Palamanit, ***3.Anil Kumar*** and Perapong Tekasakul, Mechanical, DTU

24. Multi View Image Registration and Removal of Transformation Errors, S Kumar, R Gupta, ***3.R Kapoor***, and ***3.A Bhattacharyya,*** Electrical and Electronics, DTU

25. N gram Based Machine Translation for English Assamese: Two Languages with High Syntactical Dissimilarity, Zakir Hussain, ***6.1.Malaya Dutta Borah*** and Abdul Hannan, Computer Science, DTU

26. Optimization of FSW Process parameters for achieving a goodquality joint, ***6.Satyaveer Singh***, ***3.N. Yuvraj***, ***3.Reeta Wattal***, Mechanical, DTU

27. Parboiled Paddy Drying with Dierent Dryers: Thermodynamic and Quality Properties, Mathematical Modeling Using ANNs Assessment, Ebrahim Taghinezhad, Antoni Szumny , Mohammad Kaveh, Vali Rasooli Sharabiani, ***3.Anil Kumar*** and Naoto Shimizu, Mechanical, DTU

28. PERFORMANCE ANALYSIS OF C.I. ENGINE USING BIODIESEL FUEL BY MODIFYING INJECTION TIMING AND INJECTION PRESSURE, Ravi Shankar Kumar, ***6.Dharamveer Singh*** and Desh Bandhu Singh, Mechanical, DTU

29. Petal shaped nanoantenna for solar energy harvesting, ***6.Ritika Ranga***, ***3.Yogita Kalra*** and ***3.Kamal Kishor***, Physics, DTU

30. Pragmatic Insights: Live UIs for Agent-based Modelling in Pharo, ***6.Jigyasa Grover***, Nick Papoulias, Serge Stinckwich and Pierre Bommel, DTU

31. Prediction of Employee Performance using Machine Learning Techniques, ***3.1.Anu Singh Lather***, ***3.Ruchika Malhotra***, ***8.Priya Saloni***, ***8.Prabhjot Singh*** and ***8.Sarthak Mittal***, CSE, DTU

32. Probing bulk viscous matter-dominated model in Brans-Dicke theory, ***3.C.P. Singh*** and ***6.Simran Kaur***, Applied Mathematics, DTU

33. Reusing marble dust as reinforcement material for better mechanical performance: studies on compositing aluminum matrix, ***6.Vipin Kumar Sharma***, ***6.Sumit Chaudhary***, ***6.Vikas***, Sonia and ***3.Ramesh Chandra Singh***, Mechanical, DTU

34. Statistical Analysis for Parametric Optimization of Gas Turbine-Steam Turbine Combined Power Cycle with Different Natural Gas Combustion, Kaushalendra Kumar Dubey and ***3.R.S. Mishra,*** Mechanical, DTU

35. Temporal Probability Assessment and Its Use in Landslide Susceptibility Mapping for Eastern Bhutan, Abhirup Dikshit, ***3.Raju Sarkar***, Biswajeet Pradhan, Ratiranjan Jena, Dowchu Drukpa and Abdullah M. Alamri, Civil, DTU

36. Textile Based Chemical and Physical Sensors for Healthcare Monitoring, Amir Hatamie, Shayan Angizi, ***6.Saurabh Kumar[1]***, ***6.Chandra Mouli Pandey[2]***, Abdolreza Simchi, Magnus Willander and ***3.Bansi D. Malhotra[1]***, [1]Biotechnology and [2]Applied Chemistry, DTU

37. Utilizing Temporal Psycholinguistic Cues for Suicidal Intent Estimation, Puneet Mathur, Ramit Sawhney, ***8.Shivang Chopra*** and Rajiv Ratn Shah, Computer, DTU

1. *Chancellor*

2. *Pro Vice Chancellor*	***2.1.*** *Ex Pro Vice Chancellor*
3. *Faculty*	***3.1.*** *Ex Faculty*
4. *Teaching-cum-Research Fellow*	***4.1.*** *Asst. Librarian*
5. *Alumni*	
6. *Research Scholar*	***6.1.*** *Ex Research Scholar*
7. *PG Scholar*	***7.1.*** *Ex PG Scholar*
8. *Undergraduate Student*	***8.1.*** *Ex Undergraduate Student*

A survey on database intrusion detection: approaches, challenges and application

Rajni Jindal* and Indu Singh

Department of Computer Science and Engineering,
Delhi Technological University,
Delhi, India
Email: rajnijindal@dce.ac.in
Email: indusingh@dtu.ac.in
*Corresponding author

Abstract: Databases store vital information of an organisation and are therefore integral for its efficient working. This necessitates the establishment of database intrusion detection systems (DIDSs) which can detect and prevent unauthorised user access to the critical information stored in database. A lot of work has been done in the field of DIDSs which has grown at a very rapid pace. A large number of publications emerging every year to further improve upon the existing state of the art solutions. This paper investigates research on major approaches proposed in the field of database intrusion detection and analyses the drawbacks of the proposed methods in order to drive future research towards more efficient and effective DIDSs. A systematic survey is conducted in order to classify various approaches for detecting intrusion in databases. The work identifies open research questions and challenges, by methodically comparing existing strategies to combat malicious transactions in a database system, and also provides an insight to the applications of DIDSs.

Keywords: database security; database intrusion detection; insider attack; role-based access control; data dependency mining.

Reference to this paper should be made as follows: Jindal, R. and Singh, I. (2019) 'A survey on database intrusion detection: approaches, challenges and application', *Int. J. Intelligent Engineering Informatics*, Vol. 7, No. 6, pp.559–592.

Biographical notes: Rajni Jindal is currently heading the Department of Computer Science and Engineering at Delhi Technological University. She joined Indira Gandhi Technical University for Women (IGDTUW) as a Professor in 2012 on lien. She worked as the Head (IT) and Dean (Research and Collaboration) at IGDTUW till February 2015. Her major areas are database systems, data mining and operating systems. She has supervised above 55 MTech thesis and authored/co-authored around 80 research papers. There are 12 PhD students working under her supervision and four PhD have already been awarded. She has authored books on *Data Structures Using C and Compiler-Construction and Design*. She has successfully organised IEEE International Conferences ICDMIC-2014 and IICIP 2016.

Indu Singh is an Assistant Professor in Computer Science Engineering Department at Delhi Technological University, Delhi. Singh holds a BTech in Computer Science Engineering and an MTech degree in Information Security from Ambedkar Institute of Advanced Communication Technologies and Research, Guru Gobind Singh Indraprastha University, Govt. of NCT Delhi. Currently, she is pursuing her PhD in Computer Science and Engineering,

Copyright © 2019 Inderscience Enterprises Ltd.

specialising in Data Mining and Information Security at CSE Department, Delhi Technological University, Delhi. Her research interests include database systems, data mining, information security and machine learning. She has published papers in international conferences and journals in IEEE, Elsevier, Springer and ACM. She has also served as a reviewer for several International conferences of IEEE and Springer in India and abroad.

1 Introduction

Databases play an indispensable role in the smooth working of an organisation. Databases are responsible for storing crucial enterprise data, and are therefore subjected to a number of threats. Hence, it is integral that Databases are defended against attempts of intrusive access. Databases are exposed to two different types of attacks – external and internal. While, it is important to safeguard databases against unauthorised access by sources which are not a part of the organisation, it is also imperative that the sensitive information is protected from employees exploiting their access privileges. Intrusion detection systems (IDS) detect attacks against information systems by monitoring their usage and detecting the presence of insecure states. It primarily collects data from a computer system, analyses it to search for security issues, and presents the outcomes to the administrator (Bace and Mell, 2001). Debar et al. (1999) proposed one of the earliest works that introduced the taxonomy for various intrusion detection systems highlighting various aspects of this area. Intrusion detection systems can also be classified as network-based and host-based IDSs. Network-based IDSs mechanisms detect attacks by capturing and analysing network packets. Network-based IDSs consists of uni-functional hosts that are designed to seize network traffic in different parts of network and redirect attacks to single management console. Recently, Medhat et al. (2017) designed a two-layered and three-layered intrusion detection system (IDS) for wireless sensor networks (WSN) using supervised and unsupervised learning at different levels. Wankhade and Jondhale (2019) used the ensemble clustering technique for intrusion detection which increases detection rate and lowers false alarm rate. As described by Kumar et al. (2006), Host-based IDS are capable enough to work with high quality informative data. These systems examine clients' activities and behaviour on a given machine. Host-based IDSs are usually aided by highly informative and quality data as described by Nazer and Selvakumar (2011).

Database intrusion detection systems (DIDS) as elucidated by Bertino and Sandhu (2005) are implemented to detect and prevent security violations in a database and observe discrepancies in user-access patterns to recognise intrusive attempts to access sensitive information in a database. And Lundin and Jonsson (2002) advocated that DIDs are categorised into different types. For detection of known patterns of attacks, signature-based database IDSs by Denning (1987) are employed. However, they are susceptible to novel attacks and require regular updates. These real-time signature updates introduce a large overhead. Another approach for dynamic signature recognition (El Rahman, 2017) has recently considered stroke related features followed by outlier analysis which tends to enhance run-time performance for differentiating genuine and forgery signatures. In the working of non-signature database-based IDS developed by Vaccaro and Liepins (1989), the input transactions are evaluated against a known user

pattern, and the resulting divergence between the probability distributions categorises the input as intrusive or normal activity. These systems have turned out to be efficient in dealing with insider attacks.

For example, Chagarlamudi et al. (2009) proposed an approach to prevent malicious insider activities at the database application level. User tasks were implemented using Petri nets and tests were carried out to check whether transactions conformed to a pre-determined order or not. In another example, Denning (1987) underlined the importance of a real-time intrusion detection system based on the hypothesis that intrusions into the system involve abnormal usage of the system. The abnormal usage may include high rate of login failures, unusual time access, unusual data access patterns, etc. An intrusion detection expert (IDE) system is discussed as a separate entity which adds as an extra layer of security for the database.

A lot of work has been done in the field of DIDSs since their inception. The field has grown at a very rapid pace, with large number of publications emerging every year to further improve upon the existing state of the art solutions. This paper strives to underline some of the landmark approaches that have been presented so far post 2010. Before that Santos et al. (2014) have given insights for the field of DIDS. The intent is to do the following: highlight the comparisons as well as flaws of the different DIDS approaches, present a discussion on their weaknesses and strengths based on prevention or termination of user action when abnormal access is observed, look for accuracy in detection and the practical limitations of such systems and also discuss intrusion prevention system, DIDS with proactive response mechanisms. In DIDS, response plays a key role in mitigation of these attacks. Intrusion detection systems without a prompt response are not effective even if they can detect threats and generate alarm.

The rest of the paper is organised as follows. The remaining part of this section gives an introduction to database intrusion detection systems. It also discusses different types of intruders penetrating through the security systems. Section 2 enumerates the research methodology followed by us. Section 3 presents the overview of prominent approaches for database intrusion detection systems based on various parameters. Section 4 discusses the challenges faced while performing the research. Section 5 illustrates various applications of intrusion detection in databases. Section 6 discusses the significant finding and impact of the research work going on in DIDS followed by a conclusion in Section 7.

1.1 Database intrusion detection systems

Over the years, as databases evolved they became the preferred form of storing vast amounts of data. Database management systems have increasingly made accessing and concurrently modifying data, easy and efficient. These systems have also led to an increase in illicit access to the database by exploiting various vulnerabilities. Although firewalls act as a security mechanism to monitor incoming and outgoing traffic, they have been found to be under-prepared to protect the database from a wide range of attacks and therefore need to be complemented with intrusion detection systems. An intrusion is defined as any set of actions that attempt to compromise the integrity, confidentiality or availability of the resource as suggested by Sandhu et al. (2000).

Given the importance of data in an economic context, any limitation is a critical issue that might make the currently available systems unsuitable or even incompetent solutions for many database systems. Hence the development of an intrusion detection system which defends the system against illegal access is crucial. There are two major approaches to conquer the problem of detecting intrusions in a database. One is to check the queries against well-known attack patterns (misuse detection), and another way is to check the deviation of observed behaviour from 'normal' or legit user behaviour (anomaly detection).

1.2 Types of intruders

Intruders attempt to penetrate through security systems of organisations in a number of ways. This may involve misuse of privileges or false identity operations. Intruders are classified into separate categories each with different level of vulnerabilities as expressed by Anderson (1980) as follows:-

1 legitimate user: when an authorised user who has regular access to the database uses the records with a malicious intent

2 clandestine user: when the user clutches supervisory control of the system and operates below the level of action of audit trails

3 masquerader: it is someone who pretends to impersonate another user in the system and uses its privileges to illegally access the data.

4 external intruder: when a user who does not have access to the database bypasses the security of the database through loopholes in the system or SQL injections.

1.3 Types of attacks

According to intruders' intentions there are mainly three types of attacks as follows (Burtescu, 2009):

1 Attacks aimed at compromising integrity: in these types of attack, the intruder seeks to compromise the assurance of authorised modification. Thus, the attacker aims to modify data reliability and accuracy, such as corrupting or deleting the data in a given database object, abuse of access/malicious use, internal attacks, etc. (e.g., masquerading and modification).

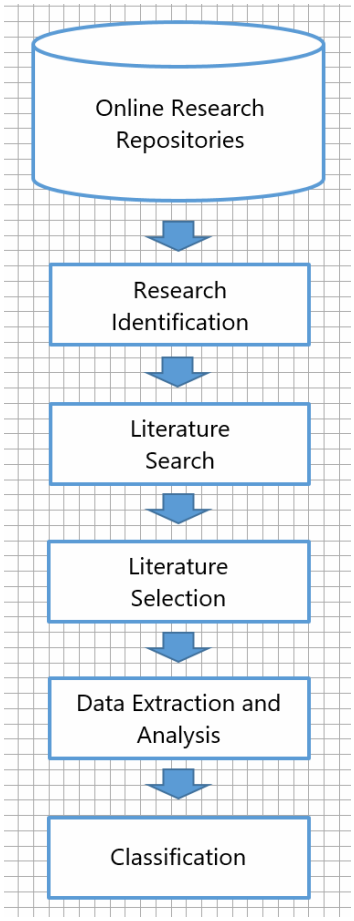
2 Attacks aiming at compromising confidentiality: in these attacks, the intruder focuses on gaining unauthorised access, such as stealing business information, rather than damaging data, attempted break-ins, data leaking, etc. (e.g., snooping).

3 Attacks aiming at making the database unavailable (availability attacks). These attacks aim on making database services inaccessible to authorised users, i.e., they are mainly denial of service (DoS) attacks (e.g., flooding database services and bandwidth with a large number of requests, crashing database server instances, deleting database objects, etc.).

2 Research methodology

In this survey, a systematic literature review is carried out along the lines followed by Ferraiolo et al. (1992) and Kitchenham (2004), which was based on research synthesis procedures (e.g., Cooper et al., 2009). First research questions are defined and then an extensive literature search to select literature is carried out. The resulting data set was used in synthesising the literature based on their underlying techniques and security concepts. Further, these concepts are classified to identify research challenges and gaps. The research methodology of this paper is outlined in Figure 1.

Figure 1 Research methodology (see online version for colours)



2.1 Research identification

In this paper, the attempt is to study and evaluate the progress of information security techniques specific to intrusion detection systems for databases. This goal is accomplished by discussing answers to the following questions:

Q1 What has been the usual methodology for handling security for the wide gamut of intrusions and database security breaches?

Q2 Are the methodologies used in conjunction with each other or alone and how effective have been these methods for different types of intrusions?

Q3 How the DIDS approaches have been tailored to account for dynamic data?

Q4 What has been the effect of present state of the art DIDS solutions on detection and false alarm rates?

Organisations increasingly use databases and database management systems to store vital operational data. The security of these systems that manage data therefore has become extremely crucial. After carrying out an extensive study of the current methods used in the field of database intrusion detection systems, significant work carried out in the domain of database security has been uncovered database IDSs come across a wide range of attacks and attempts to breach security mechanisms in place.

The first question was helpful in understanding what has been the usual methodology for handling security in case of various types of intrusions and data security breaches. Having understood the underlying structure of intrusion detection approaches Q2 evaluates the state-of-the-art solutions to determine if methodologies can be used in conjunction with each other and how effective are these approaches for handling different types of intrusions. Thereafter, through Q3 it has been ascertained if these approaches have been tailored to account for dynamic data. Finally, by Q4 it has been found out the effects of present state-of-the-art solutions on detection and false alarm rates.

2.2 Literature search

The literature search consisted of both a general and a specific search.

The general search involved the use of platforms such as Google Scholar (http://scholar.google.com), IEEE Computer Society (http://www.computer.org) and the online IEEE Xplore Library (http://ieeexplore.ieee.org), ACM Digital Library (http://www.acm.org) and Springer (https://link.springer.com/). On all the aforementioned platforms the searches were filtered using the keywords mentioned at the beginning of the text.

To perform specific literature searches, top Journals and Conferences were selected as mentioned in Table 1. This selection was based on field-specific citation-based journal ranking metrics such as impact factor. Upon examination if the literature was found to follow the set keyword filters, then it was marked to be potentially pertinent. Altogether 163 potentially pertinent publications were found after administering the literature search.

2.3 Literature selection

Having established the set of publications to examine, focus is shifted to selection of pertinent literature. Publications are scrutinised in accordance to the selection scheme described in Figure 2. First, the literature's title, its set of keywords (if explicitly mentioned in the literature), and the paper abstract were assayed.

Table 1 Significant journals and conference

Article	*Citations*	*Total (%)*	*Journal*	*Areas*
Denning (1987)	4,555	23.22	*IEEE Transactions on Software Engineering*	Abnormal behaviour, auditing, intrusions, monitoring, profiles, security, statistical measures
Vaccaro and Liepins (1989)	268	1.37	*IEEE – Symposium on Research in Security and Privacy*	Anomaly-detection, statistical approach, audit logs, clustering, rule forests
Lunt (1993)	506	2.58	*Elsevier – Computers and Security*	Fuzzy sets, possibility theory, fuzzy systems, approximate reasoning, flexible constraints.
Sandhu (1996)	276	1.41	*Computer Security – Esorics'96 (Rome) Eds., Lecture Notes in Computer Science, Vol. 1146, Springer-Verlag, New York*	Lattice-based mandatory access control, independent and composite confidentiality, integrity roles
Sandhu and Samarati (1996)	188	0.96	*ACM Computing Surveys (CSUR)*	Password-based authentication, token-based authentication, biometric-based authentication, access control, audits
Sandhu (1998)	8,470	43.18	*ACM – Advances in Computers*	Role-based access control, administrative role-based access control, role, lattice
Chung et al. (1999)	279	1.42	*Springer – Working Conference on Integrity and internal control in Information systems*	Misuse detection, privileged users, insider abuse, frequent item-sets, primary and foreign key dependency, auditing
Lee et al. (2000)	104	0.53	*ACM – Real Time Technology and Application Symposium (RTAS)*	Time semantics, real-time database systems, intrusion detection
Osborn et al. (2000)	803	4.09	*ACM Transactions on Information and System Security (TISSEC)*	Role-based access control, mandatory access control, lattice-based access control, discretionary access control
Debar and Wespi (2001)	781	3.98	*ACM – International Symposium on Recent Advances in Intrusion Detection*	Intrusion detection, alert aggregation, alert correlation, unified data model for alert handling
Lee and Xiang (2000)	678	3.46	*IEEE – Symposium on Security and Privacy*	Intrusion detection, signatures and misuse detection, anomaly detection
Liu (2002)	144	0.73	*ACM – Annual Computer Security Applications Conference*	prevention-centric security, intrusion-tolerant, attack-resilient
Hu and Panda (2004)	198	1.01	*ACM – +Symposium on Applied Computing*	Data mining, intrusion detection, database security
Bertino and Sandhu (2005)	462	2.36	*IEEE – Transactions on Dependable and Secure Computing*	Data dependency, database intrusion detection, database privilege abuse, access control
Zhong et al. (2005)	10	0.05	*IEEE – International Conference on Machine Learning and Cybernetics*	Q-cluster clustering algorithm; query similarity; intrusion detection; database security

Table 1 Significant journals and conference (continued)

Article	*Citations*	*Total (%)*	*Journal*	*Areas*
Bertino et al. (2005)	154	0.79	*ACM – Annual Computer Security Applications Conference*	RBAC, logs, intrusion detection, privacy preserving
Spalka and Lehnhardt (2005)	47	0.24	*ACM – International Conference on Recent Advances in Intrusion Detection (RAID)*	Database security, anomaly detection, misuse detection, relational databases
Hu et al. (2006)	187	0.95	*NIST Interagency Report 7316, Nat'l Institute of Standards and Technology*	Access control, authorisation, policy, computer security
Scarfone and Mell (2012)	966	4.92	*Computer Security Resource Centre (National Institute of Standards and Technology)*	Host-based and network-based intrusion detection and prevention system (IDPS), network behaviour analysis
Bishop and Gates (2008)	97	0.49	*Workshop on Cyber Security and Information Intelligence Research*	Insider threat, insider detection, access control
Motwani et al. (2008)	38	0.19	*IEEE – International Conference on Data Engineering*	Auditing, semantic suspiciousness, syntactic suspiciousness, perfect privacy
Fonseca et al. (2008)	35	0.18	*ACM – Symposium on Applied Computing*	Intrusion detection, SQL injection, audit logs
Bockermann et al. (2009)	50	0.25	*ACM: DIMVA – International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*	Outlier detection, clustering, contextual awareness, tree-kernels for correlation
Probst et al. (2010)	51	0.26	*Springer – Insider Threats in Cyber Security*	Insider threats, policy hierarchy
Mathew et al. (2010)	110	0.56	*Springer – International Conference on Recent Advances in Intrusion Detection*	User role profiling, user database access patterns, insider threats
Kindy and Pathan (2011)	66	0.34	*IEEE Symposium on Consumer Electronics (IEEE ISCE)*	SQL injection attacks, detection, countermeasures
Viet et al. (2012)	6	0.03	*IEEE International Conference on Systems, Man & Cybernetics*	Insider attack, collaborative attack, information flow, cyber-attack, information security
Kindy and Pathan (2012)	19	0.10	*Journal of Communication Networks and Information Security (IJCNIS)*	SQL injection attacks, vulnerability, web security
Nurse et al. (2014)	57	0.29	*Security and Privacy Workshops (SPW), IEEE*	Insider threat, threat framework, technical, psychological indicators, attack chain
Bertino (2013)	6	0.03	*Springer International – VLDB Workshop on Secure Data Management*	Data subjects, data users, insider threats, big data
Thuraisingham et al. (2016)	4	0.02	*IEEE – International Conference on Information Reuse and Integration (IRI)*	Data-centric, policy-based sharing, data trustworthiness, risk-based approach

Figure 2 Literature selection

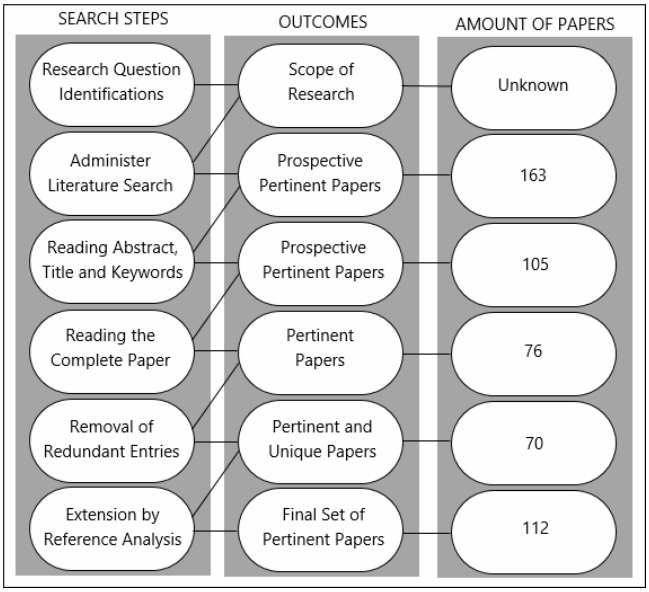
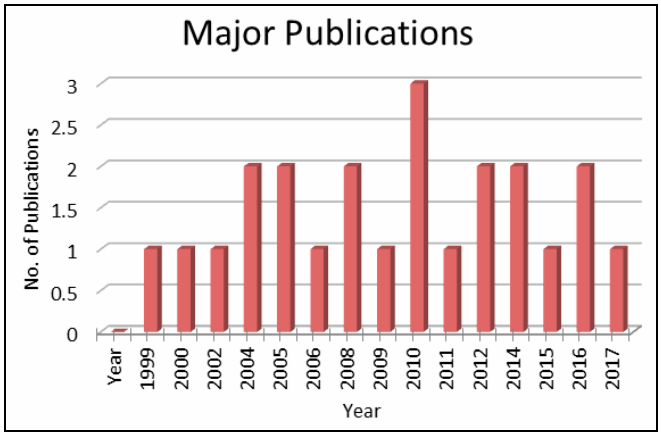


Figure 3 Number of publications (see online version for colours)



Hypothetically, one can assume that authors have used the keywords as mentioned in Table 2 or their synonyms as a part of the titles or subtitles of their papers. But in practice, authors tend to focus their titles to specific areas of research. This came across as a major hurdle in the selection process, which was overcome by evaluating the complete literature. This was the second phase of the selection process. Upon evaluation if the literature was found to be related to the scope of study, then the publication was classified as pertinent literature.

For example, publications in the field of 'access control', 'association rule mining' were classified as pertinent. Finally, as an additional filtering and classification phase, the references of selected literature were assayed to reveal potentially pertinent publications and to establish the relevance of the literature under consideration. This process resulted

in a set of 105 publications from journals and proceedings of conferences. The potential problem of duplicate publications, i.e., articles published in both journals and conferences is solved by preferring journals which extended on the conference literature.

As mentioned earlier, reference assay was used to reveal potentially relevant publications, and to uncover literature which may not have been indexed in online searches. It should also be noted that some publications may not have been available in online libraries, thus it is possible that the most recent publications were missed. Figure 3 illustrates the major publications from 1999 to 2017. The set of all major journals and conferences can be found in Table 1.

Table 2 Major set of keywords

	CONCEPTS USED
Data mining	Sequential pattern mining, weighted sequential pattern mining, frequent pattern mining, constraint based FPM, statistical data mining, association rule mining, clustering, temporal analysis, knowledge discovery
	SECURITY CONTROLS ATTRIBUTES
Intrusion detection	anomaly detection, misuse detection, command based, statistics based, alert aggregation, host-based, network-based, signature based intrusion detection, expert system
Database security	Data confidentiality, integrity, availability, accountability, privilege abuse, malicious transactions, insider threats, cyber-attack, information security, non-repudiation
Insider threats	Authorisation abuse, confidentiality compromised, privilege abuse, collaborative attack, weak authentication, information misuse, masquerade, clandestine user, SQL injection attack, hardware misuse, witting/unwitting, data corruption, database misconfiguration
	ACCESS CONTROLS
Access control models	Role-based access controls, access levels, authorisation, user roles, user profiles, transaction profiles, rule based access control
Constraints	Authorisation constraints enforcements, policies, integrity constraints
	APPLICATIONS
Access control application	Authorisation, flexible policies, security policies, privacy preservation, data theft prevention, health management, medical sector, retail sector, military sector, education sector, banking
	ERROR HANDLING
Error detection	Anomalous access patterns, policy violation, intrusion detection, data extraction, audit logs
Error prevention	Damage assessment, regeneration, recovery

2.4 Extraction and synthesis of data

The most important challenge lied in classifying publications in a significant and expressive structure. A detailed study of the concepts in data mining such as sequential pattern mining, weighted sequential pattern mining, association rule mining and clustering was carried out.

Firstly, the publications were categorised depending on the security control attributes, including database security, insider threats. Work in database security comprised of data

confidentiality, integrity, and availability, database privilege abuse, malicious transactions, insider threats, cyber-attack, information security and non-repudiation. Insider threats dealt with authorisation abuse, privilege abuse, collaborative attack, weak authentication, information misuse, masquerades and clandestine users, SQL injection attacks and penetration. In the next stage, publications were categorised based on the access control techniques incorporated in them. It was found that significant work has been carried out in access control models including role-based access controls (RBAC), authorisation, roles, and user and transaction profile. Constraints relevant to Access Control were also incorporated in the study, thereby including authorisation constraints, enforcement constraints and various policies.

Finally, the publications were grouped according to their applications, verification and error handling capabilities. The error detection module focuses on work carried out in the field of anomalous access patterns, policy violation, binding, intrusion detection, data extraction and audit logs. Error prevention module laid stress upon damage assessment, regeneration and recovery.

2.5 Classification of security controls

Once the main areas for research were established, the next goal was to assess the results to identify the scope of the research, i.e., the research challenges that are to be explored. The classification involved the study of the DIDS approaches, the post detection action, performance metrics and system restrictions.

Thus, the DIDS were categorised according to their underlying principle techniques. Additionally, the system action after detection was used as another classification filter. This classification helped in the identification and documentation of the research challenges (Section 4) in DIDS. The classification scheme can also be based on higher-level aspects such as user privacy in DIDS systems, attack classification by the system and access control methods used in system.

This can be summarised into a four-step pipeline consisting of the following steps:

1 define and document the scope and goals of the study by defining research questions that are to be answered by the end of the study

2 a thorough and comprehensive literature search which involved both a general as well as a specific search

3 literature selection by picking publications pertinent to the scope of study

4 classification of the literature into buckets of the underlying techniques and specific high-level categories resulting in definition of research challenges.

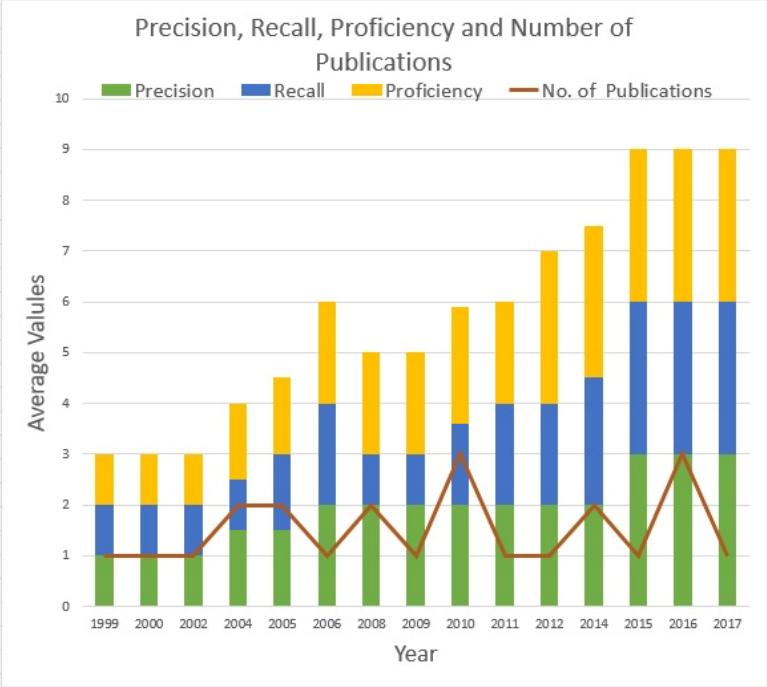
3 Literature review

In the following few sections, developments in the field of database security and DIDS are extensively reviewed, and in the process the years of publication and sources of the selected literature have been outlined. Figure 4 traces the progress of the average values of evaluation metrics such as precision, recall and proficiency from 1999 to 2017. Figure 5 illustrates the relative proportions in which such techniques such as statistical analysis; clustering, dependency analysis etc. have been used. Dependency and relation

analysis is the most widely used technique in database intrusion detection. Table 5 introduces some of the most commonly used evaluation metrics such as accuracy, reliability diagram, etc.

Data mining and knowledge discovery using data access logs have been widely used throughout the length and breadth of research in database intrusion detection. Techniques such as dependency and relation analysis, sequence alignment analysis and information flow patterns (refer Figure 5) use core knowledge discovery principles to mine user access patterns and thereby distinguish intrusions from normal accesses. In the following sections, besides statistical data analysis, some important and landmark research on database intrusion detection involving data mining and pattern discovery have been studied briefly.

Figure 4 Change in accuracy over time (see online version for colours)



3.1 Overview of selected publications

3.1.1 Dependency and relation analysis

Intrusion detection techniques that employ dependency and relation analysis use the relations and dependencies between different data items which can be represented by the order in which these items are accessed.

The DEMIDS approach was proposed by Chung et al. (1999). It derives user profiles from user access patterns in the database audit logs. The distance of a new query initiated by a user from the derived user profile is measured using distance metrics to detect an anomaly. The user profiles are established through frequent sequences, which are

determined by feature vectors. These feature vectors are based on syntactical analysis of user commands containing attributes that are issued together.

Zhong et al. (2005) organise each user query into a tuple. Constraints are used to reduce frequent sequences and improve system performance. Frequent item-sets are formed by mining user query profiles using these tuples. The algorithm uses query distance as a metric to calculate the difference on constraints in any random query and the frequent sequences.

Kamra et al. (2008) proposed a RBAC intrusion detection system as an advancement of Bertino et al. (2005) by developing three models of different granularity to represent SQL queries appearing in the database log files.

Bertino et al. (2005) introduced the concept of sensitivity of attributes, as every attribute is given a measure of importance on the basis of support in the analyse transaction. Rules generated using a weighted data mining algorithm mirrors the dependencies among database attributes. Transactions which do not follow these rules are categorised as malicious.

3.1.2 Temporal analysis

These techniques exploit the real time properties of data in intrusion detection like time span between user actions and the time interval of those actions. For capturing time semantics of data objects time signatures are applied in these techniques to detect violations of the security policy.

The proposed model (Lee et al., 2000) is constructed from time signatures by integrating standard deviation and mean in the system. The real time system works on a previously defined time span in a real time system. In the proposed approach, transactions are defined as a series of read and write queries for every data attribute performed in a specified update time-period. It also results in an efficient performance on a distorted dataset.

3.1.3 Command template analysis

Command template-based DIDS matches the raw transactional log of SQL queries against incoming database transactions efficiently. The repeated occurrence of set of SQL statements is listed according to the frequency of each learnt fingerprint .For correlating commands with applications and to differentiate between benign and malicious queries, parse tree of commands was used.

Lee et al. (2002) proposed a system in which a set of legitimate queries are summarised into SQL templates in which represents the models of all the queries. A Kolmogorov-Smirnov test, at 90% confidence level is implemented to decide whether an unrestricted variable or a finite list of values should be employed. Each learnt fingerprint's frequency is also calculated.

Bockermann et al. (2009) proposed a grammar-based analysis of frequently used vector-based data using tree-kernel based machine-learning techniques. Commands are correlated using the parse-tree structure of SQL. The changes in command syntax tree are analysed to determine the authenticity of the user-entered command. In learning phase, SVM is implemented. Clustering is used for outlier detection and determining the authenticity of user-entered commands.

3.1.4 Integrated dependency with sequence alignment analysis

Some data items may be read or written before a particular item is updated. Data items continue to be read and written after the update. These updates read and write database access operations are stored in DBMS audit trail that can be used for finding data dependency relationships. These dependency relationships are then quantified and evaluated using Sequence Alignment technique. It detects the longest common subsequence and measures the similarity of a new database access sequence with existing normal sequences and helps in anomaly detection.

In 2006, Srivastava et al. proposed an intrusion detection algorithm named weighted data dependency rule miner (WDDRM) for interpreting dependencies among the data items. The transactions that did not follow up the extracted data dependency rules were marked as malicious. Every attribute is associated with its sensitivity value in the form of weights.

In 2004, Hu and Panda proposed a method to establish dependency relationships intransaction-level attributes with high support and confidence. An update statement triggers a progression of events, i.e., updating a given attribute can result in reading or writing other attributes in the database. The model proposed a data mining approach for identifying malicious and intrusive transactions in databases. The approach is centred on extracting data dependencies between the data items present in the database. The transactions not satisfying the dependencies mined are detected as malicious transactions. The approach works effectively provided data dependencies exist in the database.

Fonseca et al. (2008) employed directed graphs to represent transactions. Transactions which do not comply with learnt execution paths are considered malicious. The command type, target objects, selected columns, and restriction attributes are used to build the execution paths. These attributes are retrieved from typical DBMS audit entries.

3.1.5 Statistical analysis

For dealing with the insider threat problem statistical analysis is used by building statistical profile of normal user's behaviour and raising an alert when a user deviates from his/her routine.

Spalka and Lehnhardt (2005) employed arithmetical functions on values retrieved from relations for detecting anomalies. Reference values include average, count, standard deviation, minimum, maximum, ranges, bit counting, zero length checking and computed ratios. Misuse is detected by inspecting objects and operations on them. Each pair of database object and operation is pre-defined as malicious or not.

Mathew et al. (2010) delineated that statistical values such as maximum, minimum, counting, median, mean, cardinality values and standard deviations are computed for all attributes of a dataset, as a result of each user command. A vector with pre-set dimensions, namely S-vector is used to store these statistics. For large datasets, S-vector is constructed using initial k or randomly picked k tuples. This maintains performance and scalability. Techniques such as naïve Bayes, clustering, support vector machines or decision trees are applied using each user's S-vector. Statistical deviation and outlier detection are implemented in detection phase to detect malicious user commands.

3.1.6 Sequence alignment analysis

Sequence alignment is a technique used to evaluate the similarity between a number of sequences. Sequence alignment can thus be applied to find whether a new database access sequence is similar to existing normal access sequences, and can therefore help in distinguishing between database access patterns of genuine users and intruders.

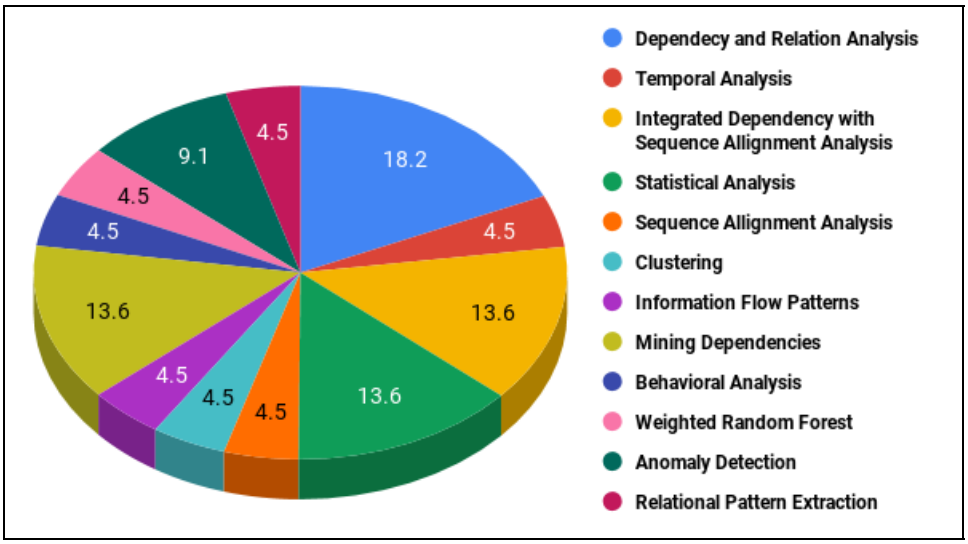
In the approach proposed by Kundu et al. (2010) user profiles are built by analysing sequences of retrieved commands, attributes and tables. The user is also provided with an option to select between organisation-based, role-based or user-based, profiles, provided the operational framework of the dataset is available. For deciding the maximum number of differences, a threshold was given through which sequence patterns were built in the learning phase. The maximum difference between the incoming sequences and the learnt ones was used to detect intrusions in the detection phase.

3.1.7 Classification

Classification is used to classify things/objects, e.g., user role profiles which helpful in distinguishing anomalous user activity from normal user activity.

In 2010, Rao et al. proposed a novel transaction level IDS to detect data privilege abuse by users of databases. The IDS is well suited for a Role Based System and implements a classifier to generate user role profiles from the normal user activity of a large number of users. The major drawback of the approach is its dependence on the order of attributes referred together in transactions and its inability to obtain correlation among various queries in a transaction.

Figure 5 Techniques used (see online version for colours)



Sohrabi et al. (2014) overcame this drawback by using leverage as rule value measure to extract infrequent but important rules. This also helped decrease the overall interesting data dependencies and increase the system performance. Rao and Singh (2014) detected the exploitation of user privileges by obtaining data dependencies between attributes in a RBAC administered database.

The major flaw of this approach is that the roles need to be defined explicitly, reducing the platform independence of the proposed system.

3.1.8 Behavioural analysis

Behavioural analysis is helpful in solving the problem of intrusion attacks due to malicious insiders by evaluating user behaviour leading to formation of groups helping in creation of routine behaviour pattern eventually leading to ease in anomaly detection.

Sun et al. (2016) illustrated a data driven evaluation model for detecting malicious insiders. They used both parallel and incremental views along with probabilistic models for auditing user behaviours. They examined anomalies from diachronic view for comparing a user's behaviour with previous data of the user.

3.1.9 Relational pattern extraction and anomaly detection

Community-based anomaly detection system was introduced by Chen and Malin (2011) for detecting malicious insider threats. It is based on the assumption that users form community structures, and users at a significant distance from the community, and having low affinity to the community are indicative of anomalous behaviour. The two major components used were relational pattern extraction along with combination of graph-based modelling, dimensionality reduction techniques and anomaly detection with nearest neighbour and clustering model. In the relational pattern extraction module, CADS deduces the community structures from CIS access logs, and obtains communities which form the core pattern. Thereafter, CADS uses certain statistical formulations to measure the deviation of users from core pattern of communities to see if the users are anomalous.

3.1.10 Information flow patterns and collaborative attacks

To address the problem of collaborative insider attacks where two or more insiders work together in order to compromise important data in the information systems, illegal information's flow is used to extract relations among the system components which are fed to the detection algorithm.

Viet et al. (2012) proposed a probabilistic model to tackle collaborative insider attacks by observing the bridge data items and mutual record access probability of a data item. If a hot item, i.e., a mutually accessed data item has probability less than a specified threshold, the access is marked as a normal occurrence; else it is flagged as an anomaly.

3.1.11 Relational and statistical analysis

Chen and Malin (2011) later improvised the community anomaly detection systems (CADS) for detecting anomalous insiders. CADS used unsupervised learning scheme to identify insider threats using the access logs of community environments. CADS consisted of relational pattern extraction and deviation detection which used k-nearest

neighbour model to measure inconsistency. CADS was extended into MetaCADS which was more efficient when users were low. But MetaCADS was sensitive to number of assumed users and the number of subjects. Another flaw is that it uses an equivalent k for each user, though it is known that size of networks is variable. CADS is also prone to mimicry attacks where an attacker imitates collaborative group behaviour or the behaviour of another user.

3.1.12 Anomaly detection

An anomaly-based intrusion detection system is an intrusion detection system that generates a normal user profile and any transaction that deviates from this normal behaviour is rendered malicious. The classification employs heuristics or rules, rather than patterns or signatures to capture the normal behaviour of the users.

Hussain et al. (2015) proposed an IDS that worked on anomaly detection mechanism. They tried to detect any diversion from regular data access by applications. This was done at a transaction level. In the first phase, profiles of applications are created. The profile of every application consists of a record of all benign queries along with their constraints. This phase is called the profile generation phase. The next phase is the anomaly detection phase in which the test query is matched with the profile of the given application. If a match is found, the query is allowed else an alarm is raised.

Sallam et al. (2016) proposed a technique for detecting anomalies in user accesses by analysing profiles of normal access patterns based upon syntactic and semantic features of past and historic user queries stored in database logs. Any deviation from previous patterns is considered an anomaly. Naive Bayesian classifier was used in RBAC models to detect anomalies. When RBAC is not used, anomaly is detected using the COBWEB clustering algorithm. When there is significant overlap between access patterns, multi-labelling classifier (MLC) is used.

Sallam and Bertino (2017) proposed real time techniques for keeping a track on users' access frequencies and detecting anomalies in reference to insider threats in databases. User access profiles were generated with the help of past access logs in three stages. First stage includes the standard periodicity detection algorithm. Determining the interval of time when each of the periodic query is likely to be issued constitutes the second stage, while the third stage involves identifying relationships between queries that are generally issued together. These access profiles are then used on subsequent users' accesses for the purpose of anomaly detection.

Sallam et al. (2015) proposed an anomaly detection design particularly for RDBMS. The profiles of users which were used to detect anomalies were based on their interaction with the monitored database during the training phase. They used naïve-based classifier to search inconsistency. One of the major flaws in this design occurs when users submit similar queries within different roles. The MAP rule does not produce accurate results if the same queries are submitted during detection, since the probability that any of these queries will occur gets biased towards the roles that submit the queries.

Aljawarneh et al. (2018) gave a hybrid approach to detect network intrusions. Two main components are involved – the first component filters the data using vote algorithm with information gain and the second component applies a multitude of classifiers like – AdaBoost, random tree, naïve Bayes, etc.

3.1.13 Weighted random forest

Random forest described by Chen et al. (2004) is an ensemble machine learning technique involving unpruned classification or regression trees. These trees are generated bootstrap samples of the data by randomly selecting a subset of features from the data. Random forests may suffer from the problem of learning from extremely imbalanced data. Since these random forests are constructed in order to reduce error, these generally tend to be biased towards predicting the majority class. This results in poor accuracy with the minority class.

Ronao and Cho (2016) proposed an IDS especially customised for RBAC-administered databases. Their approach exploits random forest with weighted voting (WRF) and principal components analysis (PCA) as a feature selection technique for detecting database access anomalies. PCA produces uncorrelated and relevant features along with reduced data dimensionality. Non-uniform database query access is administered by RF as it exploits the intrinsic tree-structure syntax of SQL queries.

3.2 Intrusion response and prevention

Intrusion response and prevention may be described as the capability of ceasing intrusions both at the time of their occurrence and also before that. Intrusion response aims at handling an attack in a manner such that the damage caused to the system is minimised. Lee et al. (2000) presented a temporal analysis methodology to detect queries that demand execution outside a pre-described time schedule and prevent their execution to prevent intrusion. Sequence analysis methodology introduced by Kundu et al. (2010) prevents intrusions by withholding user activity once a suspicious set of actions is detected. The methodology however requires a considerable number of actions which make up the set, implying that an intrusion will be detected only after execution of some actions. Solutions constructed upon relational and dependency analysis as described in Bertino and Sandhu (2005), Chung et al. (1999), Zhong et al. (2005), Kamra et al. (2008) and Bertino et al. (2005) are completely capable of implementing prevention of intrusions, as each individual user command is checked and if any of them is found to be suspicious, the execution of their queries is stopped. Therefore, solutions merging a mixture of sequence and dependency analysis are adequate in performing partial prevention of intrusion.

Kamra et al. (2008) proposed an intrusion detection and prevention mechanism, improving upon the approach they proposed in 2005. They deployed policy matching to deal with the potential intrusions and clearly defined database response procedures. Kamra proposed a collection of rules syntactically similar to SQL. These were of the form: on [Event] if [Condition] then [Action] CONFIRM [Action for confirmation] on FAILURE [Failure Action] on SUCCESS [Resolution Action]. These rules enabled the security staff to describe the policies and determine the type of actions that the DIDS must take against intrusions.

Table 3 Evaluation of various prominent approaches of database intrusion detection system

Reference	*Technique*	*RBAC used*	*Threat level*	*Detection level*	*Accuracy*			*Major flaws*
					Proficiency	*Recall*	*Precision*	
Chung et al. (1999)	Dependency and relation analysis	No	External and user threats	OS level	Low	Low	Low	• Distance measure among attribute values not defined. • User profiles considered but role profiles are not considered.
Lee et al. (2000)	Temporal analysis	No	External threats towards illegal access of data	Transaction level	Low	Low	Low	• Assumes a single fixed rate for time updation.
Lee et al. (2002)	Command template analysis	No	External and user threats	Application level	Low	Low	Low	• Model is not adaptive, as it cannot dynamically adjust to different length based on run time information.
Zhong et al. (2005)	Dependency and relation analysis	No	External and user threats	Transaction level	Medium	Low	Low	• Low accuracy and low adaptability to user trends
Hu and Panda (2004)	Integrated dependency with sequence alignment analysis	No	External and user threats towards corrupting data	Transaction level	Low	Medium	Medium	• Works only when data dependencies exist. • Emphasises on malicious modification of data, while malicious read operation cannot be detected.
Bertino et al. (2005)	Dependency and relation analysis	Yes	External and user threats	Transaction level	Medium	Medium	Medium	• Does not cover transaction management or semantic integrity.
Spalka and Lehnhardt (2005)	Statistical analysis	No	User threats	Transaction level	Low	Low	Low	• Data mining techniques in analysis of group of attributes absent.
Srivastava et al. (2006)	Integrated dependency with sequence alignment analysis	No	External and user threats towards corrupting data	Transaction level	Medium	Medium	Medium	• Only update sequences are considered for malicious checks. • Poses a problem in selection of appropriate value of support and confidence.
Fonseca et al. (2008)	Integrated dependency with sequence alignment analysis	No	External threats towards illegal access of data	Transaction level	Medium	Medium	Medium	• Low accuracy and low adaptability to user trends. • Dependencies need to be maintained.
Kamra et al. (2008)	Dependency and relation analysis	Yes	Insider through RBAC, external through clustering	Transaction level	Medium	Medium	Medium	• Scalability issues considering that f-quiplet takes enough memory space for each log entry. • Query semantics have not been taken into account during the generation of quiplets.

Table 3 Evaluation of various prominent approaches of database intrusion detection system (continued)

Reference	*Technique*	*RBAC used*	*Threat level*	*Detection level*	*Accuracy*			*Major flaws*
					Proficiency	*Recall*	*Precision*	
Bockermann et al. (2009)	Command template analysis	No	External and user threats	Database level	Medium	Low	Medium	• Derivation of application specific SQL policies.
Kundu et al. (2010)	Sequence alignment analysis	No	External and user threats	Transaction level	Medium	Medium	Medium	• Detailed transaction query leads to increase in the number of false positives
Mathew et al. (2010)	Statistical analysis	No	Insider	Transaction level	Low	Low	Medium	• Computationally slow if query results are very large. • Assumes the database is stable i.e. no updates.
Rao et al. (2010)	Classification	Yes	External and insider threats given, roles are defined	Transaction level	High	Medium	High	• Binary representation of information does not scale well as historical data gets longer. Model accuracy reduces with smaller data. • The number of roles needs to be pre-defined. Role profiles are not created automatically. • Does not detect malicious read or write operations in individual transactions.
Chen and Malin (2011)	Relational pattern extraction and anomaly detection	No	Insider, collaborative threats	Transaction level	Medium	Medium	High	• Adjacency matrix representation inefficient for pattern extraction of potentially millions of users. • Performance is sensitive to the number of simulated users in a community.
Viet et al. (2012)	Information flow patterns and collaborative attacks	Yes	Insider	Transaction level	Medium	Medium	High	• Algorithm fails when a large majority data items in the data filed are updated masking malicious write operations.
Chen et al. (2012)	Relational and statistical analysis	No	Insider, collaborative threats	Transaction level	Medium	High	High	• Collaborative attacks by large groups might remain undetected as new social structures might evolve.
Doroudian and Shahriari (2014)	Mining dependencies	No	External and user threats	Transaction level	Medium	High	High	• Algorithm assumes that only a finite set of transactions issue requests to database. • Experimental determination of appropriate settings for minimum support and related constraints which does not necessarily lead to strong data dependencies.

Table 3 Evaluation of various prominent approaches of database intrusion detection system (continued)

Reference	*Technique*	*RBAC used*	*Threat level*	*Detection level*	*Accuracy*			*Major flaws*
					Proficiency	*Recall*	*Precision*	
Sohrabi et al. (2014)	Mining dependencies	No	External and user threats	OS Level	Medium	Medium	High	• Detects only malicious write operations. • Does not consider abnormal patterns in the update of the data items.
Hussain et al. (2015)	Application anomaly detection	No	Insider	Transaction level	Medium	High	High	• The generation of application profiles is static and the system cannot handle systems with dynamically changing query types. • -Considerably high FPR.
Rao and Singh (2014)	Mining dependencies	Yes	External and insider threats given, roles are defined	Transaction level	High	High	High	• Multiple passes over database to generate rules leading to computational complexity. • Roles need to be explicitly defined.
Ronao and Cho (2016)	Weighted random forest	Yes	External and user threats	Root level	High	High	High	• Limited to insider attacks at the transaction level. Does not consider collaborative attacks. • No explicitly defined mechanism to prevent malicious read operations.
Sun et al. (2016)	Behavioural analysis	Yes	Insider	Transaction level	Medium	High	High	• In the anomaly measurement, the results highly rely on selection of the parameter k.
Sallam et al. (2015)	Anomaly detection	Yes	Insider	RDBMS level	High	High	High	• Semantic features of the queries has not been taken into account for the profile generation • Attribute independency has been assumed which is not the case in real databases. • Fails when common roles have different access patterns
Sallam and Bertino (2017)	Anomaly detection	Yes	Temporal insider threats	RDBMS level	High	High	High	• No provision for recovery or prevention. • User profile updation not incorporated. • Restricted application

Table 4 Evaluation of various proposed approaches based on performance parameters

Author	*Intrusion detection capabilities*	*Detection paradigm*	*Approach for query evaluation*	*Data sensitivity*	*Prevention capability*	*Provision for recovery*	*Platform independence*	*Maintainability*	*Efficiency*
Chung et al. (1999)	Yes	Misuse	Syntax centric	Yes	Partial	No	No	Low	Low
Lee et al. (2000)	Yes	Anomaly	Syntax centric	No	Partial	No	No	High	Medium
Lee et al. (2002)	Yes	Anomaly	Syntax centric	Yes	No	No	Yes	Low	Low
Zhong et al. (2005)	Yes	Anomaly	Data centric	No	Yes	No	Yes	Low	Low
Hu and Panda (2004)	Partial	Anomaly	Syntax centric	No	Partial	No	Yes	High	Medium-high
Bertino et al. (2005)	Yes	Anomaly	Syntax centric	No	Partial	No	Yes	Medium	Medium-high
Spalka and Lehnhardt (2005)	Partial	Both, anomaly and misuse detection	Data centric	No	No	No	Yes	Low	Low
Srivastava et al. (2006)	Partial	Anomaly	Syntax centric	Yes	Partial	No	Yes	High	High-medium
Fonseca et al. (2008)	Partial	Anomaly	Syntax centric	No	Partial	Yes	Yes	High	Medium
Kamra et al. (2008)	Yes	Anomaly	Syntax centric	No	Yes	Yes	Yes	High	High-medium
Bockermann et al. (2009)	Yes	Anomaly	Syntax centric	No	Partial	No	Yes	Low-medium	Low-medium
Kundu et al. (2010)	Yes	Anomaly	Syntax centric	Yes	Partial	No	Yes	Medium-low	High-medium
Mathew et al. (2010)	Partial	Anomaly	Data centric	No	Yes	No	Yes	Low	Medium-high
Rao et al. (2010)	Yes	Anomaly	Syntax centric	No	No	No	Yes	High	Medium
Chen and Malin (2011)	Yes	Anomaly	Data centric	No	No	No	Yes	High	Medium
Viet et al. (2012)	Yes	Anomaly	Syntax centric	Yes	Yes	No	No	High	High-medium
Chen et al. (2012)	Yes	Anomaly	Data centric	No	No	No	Yes	Low	High
Doroudian and Shahriari (2014)	Yes	Anomaly and specification-based	Syntax centric	No	Partial	No	No	Medium	Medium-high
Sohrabi et al. (2014)	Yes	Anomaly	Data centric	Yes	No	No	No	Low-medium	High
Rao and Singh (2014)	Yes	Anomaly	Syntax centric	Yes	Partial	No	No	High	Medium
Ronao and Cho (2016)	Yes	Anomaly	Syntax centric	No	Yes	No	No	Medium-high	High
Sun et al. (2016)	Yes	Anomaly	Data centric	No	No	No	Yes	Low	Medium
Sallam et al. (2015)	Yes	Anomaly	Syntax centric	No	Yes	Yes	Yes	High	High
Sallam et al. (2016)	Yes	Anomaly	Both data and syntax centric	No	No	No	Yes	Medium	High

Table 5 Definition of metrics

Evaluation metrics	*Characteristic*
ROC curve	Able to rank positive and negative samples, independent from class distribution
Cost-lines	Evaluating error rate based on the different costs
Precision-recall curve	Evaluation and ranking each sample based on positive class
Gains chart	Identifying relations between true positive rate and profanity of positive classification
Reliability diagram	Investigating probability of true calibration of model
Attributed diagram	Identifying regions of model that degrade performance compared with reference models with constant performance
Accuracy	Rate of correct classification
Specificity	Rate of true negative classification

The abnormality attributes used as features for intrusion detection are – user ID, user role, IP address of different source's, client application, object type, schema, date and time of the user's action along with SQL commands with its attributes. Administration Model has been included in order to manage the algorithms and response policies for efficiently searching policies that match an anomalous request in database. The abnormality attributes used as features for intrusion detection are – user ID, user role, IP address of different source's, client application, object type, schema, date and time of the user's action along with SQL commands with its attributes. Administration model has been included in order to manage the algorithms and response policies for efficiently searching policies that match an anomalous request in database. The resolutions described by Spalka and Lehnhardt (2005) and Mathew et al. (2010) construct upon statistical analysis and rely mostly on analysis of changes made in data or on their execution results and are thus incapable of preventing intrusions. This implies that an intrusion can be detected only after an attack.

However, Spalka and Lehnhardt (2005) proposed a method which can be used to verify apriori data concerned with the rows that are to be processed. Due to this reason, information theory analysis methodology that has been presented in Lee and Xiang (2000) might accomplish partial prevention of intrusion.

A solution built upon command and template analysis as proposed in method devised by Lee et al. (2002), and Lee et al. (2002) can completely enable the prevention of the intrusion because of the same reasons as Bertino and Sandhu (2005), Chung et al. (1999) and Kamra et al. (2008).

Apart from the methodologies described previously for prevention and detection of intrusions, various other works on intrusion detection have been published. For example, the approach in by Motwani et al. (2008) presents a methodology for auditing SQL queries so as to assess their suspiciousness through a confidentiality and privacy point of view. Pei et al. (2004) present a generic survey describing how mining methodologies can be applicable for detection of intrusions. A detailed survey of SQL injection is presented in from Kindy and Pathan (2012).

The temporal approach presented in by Lee et al. (2000) showcases both intrusion detection and prevention capabilities. Hu and Panda (2004) presented a highly efficient solution which could partially detect and prevent malicious transactions. While the

approach presented in from Fonseca et al. (2008) is feasible, it only provides partial intrusion detection and prevention capabilities. Sallam and Bertino (2017) mitigate temporal insider threats by tracking access frequencies providing a real-time detection framework. Bockermann et al. (2009) presented complete intrusion detection and prevention approach, having lower efficiency.

Viet et al. (2012) presented a highly efficient intrusion detection and prevention platform dependent approach. It is inferred from the comparative analysis that there lies a trade-off between platform dependence, efficiency and intrusion detection and prevention capabilities. Tables 3 to 5 represent the evaluation of the proposed approaches.

4 Research challenges

4.1 Determining appropriate caution level

DID Systems normally use thresholds to ascertain that some action is an intrusion or not characterised by their probabilities. Determining the appropriate threshold is the major concern as very low threshold can lead to high FPR and similarly very high threshold can cause high FNR. The aim ought to be that no caution should be discarded instead they can be classified using caution level ranking techniques.

4.2 Real-time intrusion detection

The issue with most DIDS is that they detect a threat a-posteriori, i.e., after intrusion has occurred so it presents a serious challenge given the value of data for organizations. That is why a DIDS should be able to detect, react and preclude in real-time.

4.3 Damage evaluation

Most DIDS are not able to assess the ramifications of an attack on the database. Considering the importance of databases for various organisations, a decisive issue is to quickly deal with dangerous insider threats to security. So DIDS should have the capability to measure or estimate the potential impact of such an invasion. This will allow classifying invasions as tolerable or critical.

4.4 Attack type assessment

Attacks can be categorised into four major classes:

1 Snooping: snooping refers to illegal access of private or confidential information which can be used for personal corporate benefits. Snooping may also involve blackmailing.

2 Modification: it is simply changing the databases state which may cause loss of integrity and moreover could be performed in non-obvious ways making it difficult to find a trail.

3 Masquerading: masquerading or spoofing occurs when the attacker impersonates to be an authorised user in order to gain much more privileges than he is authorised for. The attacker basically fools the target system or network into validating it as a genuine device. As a result the attacker gets illegal access to sensitive information in the databases.

4 Denial of service (DoS): DoS attack is used to make a service completely inaccessible to the users including the attacker. The purpose of a DoS is to shut down the target by flooding it with a large number of request, thereby causing huge losses. Impact on performance of web server architecture is investigated using DDoS attack strategies (Singh et al., 2017).

Therefore, determining the type of attack beforehand and acting accordingly should be a major feature of the Database Intrusion Detection Systems.

4.5 ID efficiency progression

Any DIDS should have the capability to improve its performance over the course of time learning from FPR and FNR outcomes. Techniques that allow to progressively adjust the ID parameters to increase efficiency are very important and helpful.

4.6 Data burden capacity

It is important to know the limitations of the number of users and the amount of data the IDS can handle efficiently without crashing the main server.

4.7 Evolution of general purpose IDS

Most of the IDS depend on the features of the database so there is a need to develop a general purpose IDS which is independent of the database attributes on which it is applied.

4.8 Lack of benchmark datasets for DIDS

It is well known that performing an actual analysis of IDS techniques would help verify the work carried out and concluding certain facts about it like performance, efficiency, etc. But this requires certain benchmark dataset as prerequisites to be available corresponding to which this evaluation can be performed. For the DIDS techniques there are very less amount of datasets available. Mostly authors have used TPC-C, TPC-E and KDD-99 with latter being used rarely. Otherwise, authors have generally used synthetically generated workloads and datasets to resemble real world scenarios which make it difficult to generalise the performance expectation from any type of IDS.

5 Applications of intrusion-detection in databases

Applications of intrusion detection systems vary depending on the sectors where they are applied.

5.1 Military sector

Databases are utilised to store and manipulate humongous amounts of information. Characteristic of stored information varies amongst various companies and organisations.

Military databases store highly confidential information that requires top notch security. Advancements in the storage management have also brought about an urgent need to protect databases imperative. Information in a military database must not be accessed by an outsider (outsider attack) or by an unauthorised official (inside attack) because leakage of sensitive information in such cases can be perilous and life threatening. It is integral that consistency in platform, syntactic and semantic consistency along with interoperability and data storage is ensured in military databases.

Intrusion detection systems coupled with Intrusion-resilient DBMS can be utilised for protecting the database against intrusions and also for restoring its consistency in case of malicious attacks. Capabilities of systems must be enhanced by creating tools to search, transcribe, connect, assess, divide, analyse, and proceed on the right data within a small time-frame as possible. For example, partnership, inspection and decision support tools; foreign language tools; pattern analysis tools; and predictive (or anticipatory) modelling tools, are some of the core areas where the capabilities of the system must be enhanced. Desirable aspects of the system include adaptable middleware for data location, transparency and precariousness management, semantically relevant querying tuned for knowledge finding and monitoring, scalability and meditation, schema development and metadata management, and ordering unstructured data, thus preserving all the effects of legitimate transactions executed in between intrusions and their detection as reported (Popp et al., 2004)

5.2 Medical sector

The traditional file/folder concept of processing of data in the healthcare industry has been overshadowed by convenient database management systems. Amongst the various types of databases, the most recurrent are online transaction processing (OLTP) databases which enforce data integrity and prevent data loss.

Automated database systems can be used for syndrome detection (Mahmoud et al., 2018) by means of techniques like data mining and other ontology-based techniques. Also, biologically-inspired computing (Chifu et al., 2019) plays a key role in the e-health domain. Patients are being categorised for their risk of having Diabetes by using clustering based on evolutionary inspired methods.

The varied and complex datasets pose a considerable challenge to the management of medical services. Critical investigations lead to the discovery of fraudulent attempts to disrupt the traditional medical records. Because of the colossal amount of unprocessed data there is a need of techniques to augment and elevate the security of database to avoid malicious attempts to access/modify this data. Security of a medical database can be enhanced by implementing both mandatory and discretionary security policies of a database to effectively limit unauthorised access to the medical database.

The severity of the action taken varies from system to system depending on the degree of confidentiality of a particular data. Classification of various users depending on their requirements and user role hierarchy can effectively solve this problem.

An intrusion detection system combines the advantages of both the approaches to enhance the security. It takes due care of malicious attempts to steal data and its

unauthorised modification. This preserves its integrity without compromising the security issues both at the network and at the host level.

Electronic medical records (EMRs) provide convenient access to patient data. Since this data is often sensitive in nature, it must be ensured that it can be accessed by authorised users only. Role-based access control (RBAC) and experience based access management (EBAM) have been extensively used in order to enforce authorised access to sensitive patient data (Slevin and Macfie, 2007). Traditional RBAC model is not well-matched for managing impromptu and dynamic events like patients being shifted among wards, doctors taking second opinions from colleagues, or simply unexpected patient arrivals. Inevitably, most of these systems have anomaly mechanisms in state along with the usual role-based access control for managing these circumstances, which makes these apparatus much more suitable to use. However, with respect to security the operations of anomalies lead to added difficulty and a requirement to perform constant auditing to ascertain that the system is not misused. With an extraordinary system in place, it allows the users to overrule the normal access control mechanism since technical actions themselves cannot ensure privacy and security. This increases the demand for manual control mechanisms and realisation training for users to reduce the application anomaly mechanisms. However, scrutinising how these access control mechanisms are employed - in what circumstances, to handle what cases – may convey something about how the existing normal access control mechanisms should be altered to enhance them so that they suit the needs of the users, thereby removing or at least decreasing the use of such exceptional mechanisms. Also, it is fascinating to probe that if the audit logs contain the required information to track any misuse of such extraordinary mechanisms, or if not – what data is lacking (Zhang et al., 2011; Rostad and Edsberg, 2006).

5.3 Retail sector

High-profile attempts to breach data have been plaguing retail in previous years. It is believed by many security experts that the prominent technique of SQL injection has been a major component of such attacks. SQL injection came into heavy use after 1998. It is an attempt to breach security and exploit the weakness of a Web Application linked to the database through insertion of malicious SQL queries into a URI Stem/form field/cookie value for its execution. Processing such SQL injections by a vulnerable application results in rogue SQL statements being issued to the database for accessing, modifying or deleting the content that it will not generally have authorised access to. In worst scenarios, SQL injection can provide attacker with the complete control of the server where the database resides.

These high-profile attacks along with other database breaching techniques against the retailers is not only devastating for merchants but for consumers as well. Merchants suffer in terms of damaged reputation, consumer confidence and lost sales while customers suffer in terms of Personally Identifiable Information such as credit card details etc. being compromised.

Data breaches present a significant risk and therefore make every possible defensive avenue imperative in such a scenario. As malicious intruders are becoming increasingly well organised and trained to ape the patterns of legitimate customers in order to steal consumer data, detailed security strategy has become the key to this problem.

Companies must establish broader security systems incorporating IDS containing elements to efficiently contain a breach, preventing it from disrupting the database

structure and restoring the data to its consistent form. General RBAC models can be employed with focus on systemic automated approaches to role engineering, which may help in the effort of standardising RBAC too. The engineering process should be led by business policies and assumptions, assisted by data mining techniques for validation and risk assessment by Kuhlmann et al. (2003) Network-based defence is also critical and can be implemented by Network-based IDS to minimise the loss of consumer confidence.

5.4 Banking sector

In an online banking system with traditional banking methodologies, security is one of the primary concerns. In a standard database environment, the profile of the transactions that every user is permitted to evaluate is usually prescribed by the Database Administrator, as the database transactions are programmed in the database application code. In other words, the transactions are not makeshift sequences of SQL commands. On the contrary, database transactions are well specified sequences of commands implementing a limited set of predefined operation. Precaution is necessary to ensure that the information is transmitted in a safe and secure manner. The upcoming methods in Online banking system database security are incorporated to monitor and increase the security and integrity of the system. In a banking application, users are only allowed to do operations provided at the application interface (e.g., withdraw money, check account balance, etc.). End users are not allowed to execute commonly used commands.

It is extremely important to understand the need of securing important data that is essential to an organisation's profitability and reputation. For banking sector, protecting highly sensitive data such as bank account or credit card numbers and social security numbers against unauthorised access is crucial. This is because such data exists in multiple databases making it prone to theft.

As in other areas of study, the danger of malicious operations in databases is connected to DBMS susceptibilities and to likely security threats and attacks. The system vulnerabilities are an interior factor related to the set of such security systems available, the correct configuration of those systems, and the hidden flaws on the system implementation. On the other hand, security attacks are an exterior facet that mostly is affected by the objective and capacity of humans (hackers) to malevolently break into the system. Actually, the victory of a security attack rests on the vulnerabilities of the system (attacks are safe in a system where there are no potential problems) and problems are not risky if the system is not exposed to security attacks from Vieira and Madeira (2005).

It is important to realise the high amount of risk that is associated with the sensitive information stored within the databases through internal audits. In banking and financial organisations, it is essential to:

- safeguard sensitive financial data against data breaches and external attacks on the financial infrastructure (outsider attack)
- protect the database against the attempts made by insiders and privileged users having network authorisation that may cause compromise of sensitive data (insider attack).

All these issues can be tackled by implementing IDS wherein by providing least privilege to employees, the company can ensure that they can access appropriate amounts of data that is essential for them to perform their routine duties.

6 Discussion

In this section, the focus is on the significant findings and impact of the research work carried out in the field of database intrusion detection.

6.1 Principal findings

Intrusion detection systems are a single problem-single solution approach and cater to specific needs. It is seen that it is incredibly difficult to detect a wide gamut of intrusions and therefore two complimentary trends have emerged with their share of strengths and weaknesses. It can only be imagined how difficult would it be to design systems confirming with the one-solution-for-all-problems approach. For example, when Intrusion detection systems were first designed they were limited to a mainframe computer and its users. Host-based IDSs greatly simplified the intrusion detection task since an interaction from outside was sparse and unusual. As networks and the internet gained popularity, Network-based IDSs were designed to cater to newer kinds of vulnerabilities that soon transpired.

The following complimentary trends in intrusion detection have been studied:

1 misuse detection: the search for evidence of attacks based on knowledge of previous attacks

2 anomaly detection: the investigation for deviations from the normal state of a system.

Many intrusion detection techniques have been designed; however, most of the commercially available designs incorporate only one of the trends. Misuse detection is often easier to implement than anomaly detection systems. Besides, the false positives of Anomaly detection systems have made them inappropriate for commercial intrusion detection. There have been, however intrusion detection systems that have combined both the trends, such as SecureNet (from SecureNet Consortium).

Anomaly and misuse detection methodologies are distinct, and have their own set of advantages. There have however been very few successful IDSs which can differentiate between the various types of intrusions.

Intrusion detection systems like all data-driven techniques require a large number of verified (as secure) audit trails to model the normal state of a system. Therefore, extra care needs to be taken in the training phase, based on which IDSs generate the normal acceptable states of the database. With dynamic IDSs that adapt and model expected normal system states as newer verified transactions are committed to the database, come across a novel set of vulnerabilities. An attacker over time might feed valid transactions in the audit logs that increasingly make the IDS vulnerable to a certain attack. A useful analogy for such an attack can be constructed for banking transactions: an attacker might slowly and steadily increase spending from a particular account. The dynamic IDS would thus gradually expect a higher spending from the account. In this way the attacker can eventually spend a lot more from the account in effect fooling the system into thinking the transactions to be valid. Misuse-detectors are often based on signatures, and therefore fail to detect well-known attacks when these are masked and modified with a little variation. Therefore, both dynamic as well as static techniques have their share of problems. In many such circumstances Intrusion Prevention Systems in comparison to IDSs can effectively identify and prevent intrusions.

As far as detection and false alarm rates are concerned, many authors have started to focus more on them and have tried to increase detection rates while decreasing rates of false alarms. While increasing detection rates helps to mitigate threats which might pose danger to the system, decreasing false alarm rates will ensure that useful system resources are not wasted on unnecessary detections thus hampering the performance of the system. Both these help to contribute to the efficiency of the database intrusion detection deployed.

6.2 Limitations of this review

This review probes database intrusion detection systems (DIDS) and how these systems are suitable for amending security prospect of databases for better. Hence, a methodical literature review has been conducted (refer to Section 2).As there are huge number of publications pivoted on security and DIDS, literature which had major ideas from the areas of security and DIDS were chosen so as to contract down the results. Literature on security and DIDS that did not explore involvement of DIDS (e.g., implementation) was excluded. Vice versa, DIDS literature that did not accommodate any security concepts (e.g., confidentiality, security requirements) was also excluded.

7 Conclusions

This survey presents the techniques, approaches, applications and challenges of database security, in particular, database intrusion detection systems. Sectors such as healthcare, military, banking, and retail use database systems with distinctive user and processing requirements. Therefore, they require distinct Intrusion detection mechanisms. Primary difficulties in these mechanisms are: profiling user behaviours accurately, managing the false alarms generated by IDS mechanisms, increase the reliability of correlation techniques by making sure intrusions are detected and prevented before such attacks happen and damage highly sensitive data, and make sure the use of DIDS on very sensitive data environments such as that of data warehouse are credible and feasible. More importantly, it is essential that the research community and the industry should propose and develop DIDS benchmarks based on the database structure they have adopted. They should be able to discuss the objectivity and productiveness for every proposal. It can be contemplated that the challenges presented in this paper are notable and should provide motivation to fellow researchers and workers to discuss and improve upon current DIDS and possibly develop a standard benchmark to fulfil this purpose.

There is large amount of potential work left to be done in the field of database intrusion detection techniques. The possible works include determining database constraints for a given architecture to enhance the accuracy and completeness of the model, refine the signature generation scheme by including various constants, variables relational or logical operators in queries so as to improve detection accuracy and tweak existing mechanisms to incorporate static as well as dynamic aspect of analysis. Also it includes determining methods to ascertain the efficacy of the models.

References

Aljawarneh, S., Aldwairi, M. and Yassein, M.B. (2018) 'Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model', *Journal of Computational Science*, Vol. 25, pp.152–160.

Anderson, J.P. (1980) *Computer Security Threat Monitoring and Surveillance*, Technical report, James P. Anderson Company.

Bace, R. and Mell, P. (2001) *NIST Special Publication on Intrusion Detection Systems*, Booz-Allen and Hamilton Inc, Mclean, VA.

Bertino, E. and Sandhu, R. (2005) 'Database security-concepts, approaches, and challenges', *IEEE Transactions on Dependable and Secure Computing*, No. 1, pp.2–19.

Bertino, E. (2013) 'Data security-challenges and research opportunities', *Workshop on Secure Data Management*, August, pp.9–13, Springer, Cham.

Bertino, E., Terzi, E., Kamra, A. and Vakali, A. (2005) 'Intrusion detection in RBAC-administered databases', *21st Annual Computer Security Applications Conference (ACSAC'05)*, December, p.10, IEEE.

Bishop, M. and Gates, C. (2008) 'Defining the insider threat', in *Proc. of the 4th Annual Workshop on Cyber Security and Information Intelligence Research*, Tennessee, Vol. 288.

Bockermann, C., Apel, M. and Meier, M. (2009) 'Learning SQL for database intrusion detection using context-sensitive modelling', *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, July, pp.196–205, Springer, Berlin, Heidelberg.

Burtescu, E. (2009) 'Database security-attacks and control methods', *Journal of Applied Quantitative Methods*, Vol. 4, No. 4, pp.449–454.

Chagarlamudi, M., Panda, B. and Hu, Y. (2009) 'Insider threat in database systems: preventing malicious users' activities in databases', *2009 Sixth International Conference on Information Technology: New Generations*, April, pp.1616–1620, IEEE.

Chen, C., Liaw, A. and Breiman, L. (2004) *Using Random Forest to Learn Imbalanced Data*, pp.1–12, University of California, Berkeley.

Chen, Y. and Malin, B. (2011) 'Detection of anomalous insiders in collaborative environments via relational analysis of access logs', *Proceedings of the First ACM Conference on Data and Application Security And Privacy*, February, pp.63–74, ACM.

Chen, Y., Nyemba, S. and Malin, B. (2012) 'Detecting anomalous insiders in collaborative information systems', *IEEE Transactions on Dependable and Secure Computing*, Vol. 9, No. 3, pp.332–344.

Chifu, V.R., Chifu, E.S., Pop, C.B., Salomie, I. and Lupu, M. (2019) 'Evolutionary-based method for risk stratification of diabetic patients', *International Journal of Intelligent Engineering Informatics*, Vol. 7, No. 1, pp.37–60.

Chung, C.Y., Gertz, M. and Levitt, K. (1999) 'DEMIDS: a misuse detection system for database systems', *Working Conference on Integrity and Internal Control in Information Systems*, November, pp.159–178, Springer, Boston, MA.

Cooper, H., Hedges, L.V. and Valentine, J.C. (Eds.) (2009) *The Handbook of Research Synthesis and Meta-Analysis*, Russell Sage Foundation, New York.

Debar, H. and Wespi, A. (2001) 'Aggregation and correlation of intrusion-detection alerts', *International Workshop on Recent Advances in Intrusion Detection*, October, pp.85–103, Springer, Berlin, Heidelberg.

Debar, H., Dacier, M. and Wespi, A. (1999) 'Towards a taxonomy of intrusion-detection systems', *Computer Networks*, Vol. 31, No. 8, pp.805–822.

Denning, D.E. (1987) 'An intrusion-detection model', *IEEE Transactions on Software Engineering*, Vol. SE-13, No. 2, pp.222–232.

Doroudian, M. and Shahriari, H.R. (2014) 'A hybrid approach for database intrusion detection at transaction and inter-transaction levels', *2014 6th Conference on Information and Knowledge Technology (IKT)*, May, pp.1–6, IEEE.

El Rahman, S.A. (2017) 'An efficient approach for dynamic signature recognition', *International Journal of Intelligent Engineering Informatics*, Vol. 5, No. 2, pp.167–190.

Ferraiolo, D.F., Gilbert, D.M. and Lynch, N. (1992) *Assessing Federal and Commercial Information Security Needs (IT)*, No. NIST Interagency/Internal Report (NISTIR)-4976.

Fonseca, J., Vieira, M. and Madeira, H. (2008) 'Online detection of malicious data access using DBMS auditing', *Proceedings of the 2008 ACM Symposium on Applied Computing*, March, pp.1013–1020, ACM.

Hu, V.C., Ferraiolo, D. and Kuhn, D.R. (2006) *Assessment of Access Control Systems*, US Department of Commerce, National Institute of Standards and Technology.

Hu, Y. and Panda, B. (2004) 'A data mining approach for database intrusion detection', *Proceedings of the 2004 ACM Symposium on Applied Computing*, ACM, March, pp.711–716.

Hussain, S.R., Sallam, A.M. and Bertino, E. (2015) 'DetAnom: detecting anomalous database transactions by insiders', *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy*, pp.25–35, ACM.

Kamra, A., Terzi, E. and Bertino, E. (2008) 'Detecting anomalous access patterns in relational databases', *The VLDB Journal – The International Journal on Very Large Data Bases*, Vol. 17, No. 5, pp.1063–1077.

Kindy, D.A. and Pathan, A.S.K. (2011) 'A survey on SQL injection: vulnerabilities, attacks, and prevention techniques', *2011 IEEE 15th International Symposium on Consumer Electronics (ISCE)*, IEEE, June, pp.468–471.

Kitchenham, B. (2004) *Procedures for Performing Systematic Reviews*, Vol. 33, pp.1–26, Keele University, Keele, UK.

Kuhlmann, M., Shohat, D. and Schimpf, G. (2003) 'Role mining-revealing business roles for security administration using data mining technology', *Proceedings of the Eighth ACM Symposium on Access Control Models and Technologies*, June, pp.179–186, ACM.

Kumar, V., Srivastava, J. and Lazarevic, A. (Eds.) (2006) *Managing Cyber Threats: Issues, Approaches, and Challenges*, Vol. 5, Springer Science & Business Media.

Kundu, A., Sural, S. and Majumdar, A.K. (2010) 'Database intrusion detection using sequence alignment', *International Journal of Information Security*, Vol. 9, No. 3, pp.179–191.

Lee, S.Y., Low, W.L. and Wong, P.Y. (2002) 'Learning fingerprints for a database intrusion detection system', *European Symposium on Research in Computer Security*, October, pp.264–279, Springer, Berlin, Heidelberg.

Lee, V.C., Stankovic, J.A. and Son, S.H. (2000) 'Intrusion detection in real-time database systems via time signatures', *Proceedings Sixth IEEE Real-Time Technology and Applications Symposium. RTAS 2000*, pp.124–133, IEEE.

Lee, W. and Xiang, D. (2000) 'Information-theoretic measures for anomaly detection', *Proceedings 2001 IEEE Symposium on Security and Privacy. S&P 2001*, May, pp.130–143, IEEE.

Liu, P. (2002) 'Architectures for intrusion tolerant database systems', *18th Annual Computer Security Applications Conference, Proceedings*, pp.311–320, IEEE.

Lundin, E. and Jonsson, E. (2002) *Survey of Intrusion Detection Research*, Chalmers University of Technology.

Lunt, T.F. (1993) 'A survey of intrusion detection techniques', *Computers & Security*, Vol. 12, No. 4, pp.405–418.

Mahmoud, H., Abbas, E. and Fathy, I. (2018) 'Data mining and ontology-based techniques in healthcare management', *International Journal of Intelligent Engineering Informatics*, Vol. 6, No. 6, pp.509–526.

Mathew, S., Petropoulos, M., Ngo, H.Q. and Upadhyaya, S. (2010) 'A data-centric approach to insider attack detection in database systems', *International Workshop on Recent Advances in Intrusion Detection*, September, pp.382–401, Springer, Berlin, Heidelberg.

Medhat, K., Ramadan, R.A. and Talkhan, I. (2017) 'Towards distributed layered intrusion detection system for large scale wireless sensor networks', *International Journal of Intelligent Engineering Informatics*, Vol. 5, No. 1, pp.1–28.

Motwani, R., Nabar, S.U. and Thomas, D. (2008) 'Auditing SQL queries', *2008 IEEE 24th International Conference on Data Engineering*, April, pp.287–296, IEEE.

Nazer, G.M. and Selvakumar, A.A.L. (2011) 'Current intrusion detection techniques in information technology – a detailed analysis', *European Journal of Scientific Research*, Vol. 65, No. 4, pp.611–624.

Nurse, J.R., Buckley, O., Legg, P.A., Goldsmith, M., Creese, S., Wright, G.R. and Whitty, M., (2014) 'Understanding insider threat: a framework for characterising attacks', *2014 IEEE Security and Privacy Workshops*, May, pp.214–228, IEEE.

Osborn, S., Sandhu, R. and Munawer, Q. (2000) 'Configuring role-based access control to enforce mandatory and discretionary access control policies', *ACM Transactions on Information and System Security (TISSEC)*, Vol. 3, No. 2, pp.85–106.

Pei, J., Upadhyaya, S.J., Farooq, F. and Govindaraju, V. (2004) 'Data mining for intrusion detection: techniques, applications and systems', in *Proceedings of the 20th International Conference on Data Engineering*, March, pp.877–887.

Popp, R., Armour, T. and Numrych, K. (2004) 'Countering terrorism through information technology', *Communications of the ACM*, Vol. 47, No. 3, pp.36–43.

Probst, C.W., Hunker, J., Gollmann, D. and Bishop, M. (2010) 'Aspects of insider threats', *Insider Threats in Cyber Security*, pp.1–15, Springer, Boston, MA.

Rao, U.P. and Singh, N.K. (2014) 'Detection of privilege abuse in RBAC administered database', *Science and Information Conference*, August, pp.57–76, Springer, Cham.

Rao, U.P., Sahani, G.J. and Patel, D.R. (2010) 'Machine learning proposed approach for detecting database intrusions in RBAC enabled databases', *2010 Second International Conference on Computing, Communication and Networking Technologies*, July, pp.1–4, IEEE.

Ronao, C.A. and Cho, S.B. (2016) 'Anomalous query access detection in RBAC-administered databases with random forest and PCA', *Information Sciences*, Vol. 369, pp.238–250.

Rostad, L. and Edsberg, O. (2006) 'A study of access control requirements for healthcare systems based on audit trails from access logs', *2006 22nd Annual Computer Security Applications Conference (ACSAC'06)*, December, pp.175–186, IEEE.

Sallam, A. and Bertino, E. (2017) 'Detection of temporal insider threats to relational databases', *2017 IEEE 3rd International Conference on Collaboration and Internet Computing (CIC)*, October, pp.406–415, IEEE.

Sallam, A., Bertino, E., Hussain, S.R., Landers, D., Lefler, R.M. and Steiner, D. (2015) DBSAFE – an anomaly detection system to protect databases from exfiltration attempts', *IEEE Systems Journal*, Vol. 11, No. 2, pp.483–493.

Sallam, A., Fadolalkarim, D., Bertino, E. and Xiao, Q. (2016) 'Data and syntax centric anomaly detection for relational databases', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 6, No. 6, pp.231–239.

Sandhu, R. (1996) 'Role hierarchies and constraints for lattice-based access controls', *European Symposium on Research in Computer Security*, September, pp.65–79, Springer, Berlin, Heidelberg.

Sandhu, R. and Samarati, P. (1996) 'Authentication, access control, and audit', *ACM Computing Surveys (CSUR)*, Vol. 28, No. 1, pp.241–243.

Sandhu, R., Ferraiolo, D. and Kuhn, R. (2000) 'The NIST model for role-based access control: towards a unified standard', *ACM Workshop on Role-based Access Control*, Vol. 10, No. 344287.344301, pp.47–63.

Sandhu, R.S. (1998) 'Role-based access control', *Advances in Computers*, Vol. 46, pp.237–286, Elsevier.

Santos, R.J., Bernardino, J. and Vieira, M. (2014) 'Approaches and challenges in database intrusion detection', *ACM SIGMOD Record*, Vol. 43, No. 3, pp.36–47.

Scarfone, K. and Mell, P. (2012) *Guide to Intrusion Detection and Prevention Systems (IDPS)*, No. NIST Special Publication (SP) 800-94 Rev. 1 (Draft), National Institute of Standards and Technology.

Singh, K., Singh, P. and Kumar, K. (2017) 'Impact analysis of application layer DDoS attacks on web services: a simulation study', *International Journal of Intelligent Engineering Informatics*, Vol. 5, No. 1, pp.80–100.

Slevin, L.A. and Macfie, A. (2007) 'Role based access control for a medical database', *IASTED-Software Engineering and Applications Conference*, November, pp.19–21.

Sohrabi, M., Javidi, M.M. and Hashemi, S. (2014) 'Detecting intrusion transactions in database systems: a novel approach', *Journal of Intelligent Information Systems*, Vol. 42, No. 3, pp.619–644.

Spalka, A. and Lehnhardt, J. (2005) 'A comprehensive approach to anomaly detection in relational databases', *IFIP Annual Conference on Data and Applications Security and Privacy*, August, pp.207–221, Springer, Berlin, Heidelberg.

Srivastava, A., Sural, S. and Majumdar, A.K. (2006) 'Weighted intra-transactional rule mining for database intrusion detection', *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, April, pp.611–620, Springer, Berlin, Heidelberg.

Sun, Y., Xu, H., Bertino, E. and Sun, C. (2016) 'A data-driven evaluation for insider threats', *Data Science and Engineering*, Vol. 1, No. 2, pp.73–85.

Thuraisingham, B., Kantarcioglu, M., Hamlen, K., Khan, L., Finin, T., Joshi, A., Oates, T. and Bertino, E. (2016) 'A data driven approach for the science of cyber security: challenges and directions', *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*, July, pp.1–10, IEEE.

Vaccaro, H.S. and Liepins, G.E. (1989) 'Detection of anomalous computer session activity', *Proceedings. 1989 IEEE Symposium on Security and Privacy*, May, pp.280–289, IEEE.

Vieira, M. and Madeira, H. (2005) 'Detection of malicious transactions in DBMS', *11th Pacific Rim International Symposium on Dependable Computing (PRDC'05)* , December, p.8, IEEE.

Viet, K., Panda, B. and Hu, Y. (2012) 'Detecting collaborative insider attacks in information systems', *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp.502–507, IEEE.

Wankhade, K.K. and Jondhale, K.C. (2019) 'An ensemble clustering method for intrusion detection', *International Journal of Intelligent Engineering Informatics*, Vol. 7, Nos. 2–3, pp.112–140.

Zhang, W., Gunter, C.A., Liebovitz, D., Tian, J. and Malin, B. (2011) 'Role prediction using electronic medical record system audits', *AMIA Annual Symposium Proceedings*, Vol. 2011, p.858, American Medical Informatics Association.

Zhong, Y., Zhu, Z. and Qin, X.L. (2005) 'A clustering method based on data queries and its application in database intrusion detection', *2005 International Conference on Machine Learning and Cybernetics*, August, Vol. 4, pp.2096–2101, IEEE.

Materials Research Express

ACCEPTED MANUSCRIPT • OPEN ACCESS

Addition of anti-flaming agents in castor oil based rigid polyurethane foams: studies on mechanical and flammable behaviour

To cite this article before publication: Raminder Kaur Walia *et al* 2020 *Mater. Res. Express* in press https://doi.org/10.1088/2053-1591/ab68a2

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is **© YEAR The Author(s). Published by IOP Publishing Ltd**.

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 3.0 licence, this Accepted Manuscript is available for reuse under a CC BY 3.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 14.139.251.100 on 13/01/2020 at 05:33

Addition of Anti-Flaming Agents in Castor Oil based Rigid Polyurethane Foams: Studies on Mechanical and Flammable Behaviour

Raminder Kaur[a]*, Mukesh Kumar[b]

[a] Department of Polymer Science and Chemical Technology, Delhi Technological University, New Delhi-110042, India.

Corresponding Author:** ***E-mail: **rkwaliadtu@gmail.com**

[b] Department of Applied Chemistry and Polymer Technology, Delhi Technological University, New Delhi-110042, India.

Key Words: Rigid Polyurethane Foam, Castor Oil, Flame retardants, Mechanical properties, Scanning Electron Microscopy

Abstract

The major drawback in the use of Rigid Polyurethane Foam (RPUF) is its high flammability which limits is industrial applications. In the presented studies, RPUF was incorporated with different antiflamming agents such as Melamine (2,4,6-Triamino-1,3,5-triazine), Tris(1,3-dichloro-2-propylphosphate) (TDCPP) and Tris(2-chloroethyl) phosphate (TCP). Unlike the conventional RPUF, the foam developed here is derived from vegetable oil based polyol i.e. from Castor oil. Flammability of resulted RPUFs was tested according to UL 94 V test. The structure of the resulted foam samples was characterized by FTIR, XRD and SEM and their thermal properties were studied by TGA. The mechanical behaviour i.e. Tensile, compressive and flexural strength and hardness was examined by universal testing machine. The investigations made on flammability, mechanical and thermal properties of RPUFs incorporated with melamine, TCP, and TDCPP, suggested that, among all samples prepared, TDCPP

Accepted Manuscript

incorporated RPUF rendered good thermal stability, mechanical and flame retardant properties. Results obtained from SEM analysis also revealed the reduction in the cell size in case of TDCPP based RPUFs.

Abbreviations: PU-Polyurethane MDI-Methylenediphenyldiisocyanates

RPUF- Rigid Polyurethane Foam TDCPP- Tris(1,3-dichloro-2-propylphosphate)

TCP- Tris(2-chloroethyl) phosphate TGA-Thermogravimetric analysis

1. Introduction

Rigid Polyurethane Foam (RPUF) is one of the best suitable material used extensively for the production of sandwiched construction panels, railway assemblies and in buildings for insulation purposes. Due to their light weight, excellent thermal insulation, good dimensional stability, easy processing, low density and high specific strength, they are one of the most widely used material in various industrial and engineering applications including false roofing, structural panels and for the thermal insulation of buildings and refrigerators etc. Moreover, the easy processing of RPUFs make them am ideal candidate for the fabrication of many nifty products which also help to enhance the aesthetic design of homes. However, the high combustibility of RPUF makes its use restricted in many important engineering applications. As per the experimental evidences, the pores of the foam entraps air which further aid in its combustion. This cellular material burns readily in the presence of oxygen and heat with a very high fire spread rate and a high smoke release rate [1-3].The release of poisonous smoke and rapid reduction of oxygen concentration due to the combustion poses a threat to the human health and environment [4-6]. Due to fire threats of RPUFs, special fire protocols have to make compulsory by regulatory bodies for their utilization in furniture and

Accepted Manuscript

buildings [7-9].The development of the fire behaviour of the RPUF has presently been a significant issue of many studies among the researchers as it hinders its practical applications [10]. Hence, the improvement of flame retardancy of RPUFs is one of the areas of prime importance and has attracted the attention of many researchers [11, 12]. The use of fire retardants in RPUFs to decrease its combustibility and smoke production has become a vital parameter during the design and development of their products. RPUFs can be made fireproof by the incorporation of halogens, nitrogen, phosphorus, inorganic metal oxides and hydroxides based flame retardants [13]. Nitrogen and phosphorous based fire retardants are less poisonous and eco-friendly being gases themselves and produce less toxic gases or vapours products during combustion [14, 15]. In contrast, another flame retardant melamine while combustion, grips the heat of the PU matrix and go through the advanced endothermic condensation with evolution of ammonia. Polyurethane/phosphate combinations are reported to form intumescent flame retardant systems [16]. The growth of flame retardants in industrial and engineering uses is motivated due to the increased consciousness towards the safety, among the consumers, and by the environment threats posed by the researchers [14, 17-18]. Recent researches are focused on smart flame retardants such as halloysite nanotubes, which in addition to fire retardancy, also act as economic reinforcing materials for polymers [19] The present studies report the effects of different anti-flaming agents i.e. Melamine (2,4,6-Triamino-1,3,5-triazine), Tris(2-chloroethyl) phosphate (TCP) and Tris(1,3-dichloro-2-propylphosphate) (TDCPP) on the properties of RPUFs. The RPUFs developed in the given study have been derived from a plant source i.e. castor oil, the properties of which are studied and reported earlier by the some researchers [20-27] and other investigators and found to be comparable with commercial petrochemical based RPUFs.

Accepted Manuscript

2. Experimental

2.1 Raw Materials

Castor oil (AR Grade) and 4,4'-diphenylmethanediisocyanate (AR grade) were supplied by Shivathene Linopack Ltd. Parwanoo, Himachal Pardesh. Triethylenediamine (TEDA) (99%), Melamine (2,4,6-Triamino-1,3,5-triazine), Tris(2-chloroethyl)phosphate (TCP) and Tris(1,3-dichloro-2-propylphosphate) (TDCPP) were bought from Standard Chemicals (ISO 9001:2008 certified), Tilak Bazar, New Delhi. Silicon oil (C-63148-62-9) and n-Pentane (C-109-66-0) were gotten from CDH(P) Ltd., New Delhi - 110022 (India)and Glycerol (99.9%) was purchased from Sisco Industries Pvt. Ltd.

2.2 Preparation of Foams

The foam formulations are as reported in Table 1. A two-step process was used to prepare the anti-flammable RPUFs. Castor oil was modified by transesterification process in the first step to get polyol (Mpol) according to the procedure reported in prior studies [24-26]. Glycerol was used to enhance the hydroxyl value of castor oil by introducing more hydroxyl groups. The modification of the castor oil was performed under the inert atmosphere of nitrogen using 2:1 ratio of the castor oil to the glycerol at the temperature 180-210°C for 4-5 hours, till a hydroxyl value of 390-410 mg KOH/g is achieved. A schematic representation of the castor oil modification by transesterification process is as given in Figure 1. In the subsequent step, all the constituents as per the Table 1 were added to the reactor and stirred at 2200 rpm at room temperature. The resulted mixture was spread onto a metal mould (200 x 200 x 100 mm) and cured for 96 hours. On complete curing (after 96 hours), the foam was de-moulded, cut into standard dimensions and tested for its hardness, compressive, flexural and

Accepted Manuscript

tensile strength, cell morphology and density. Dimensional finishing was achieved by rubbing with fine emery papers. FTIR, XRD and TGA were performed to find out structure and thermal stability of resulted RPUFs. Fire behaviour of the resulted RPUFs was confirmed by ULV-94 test.

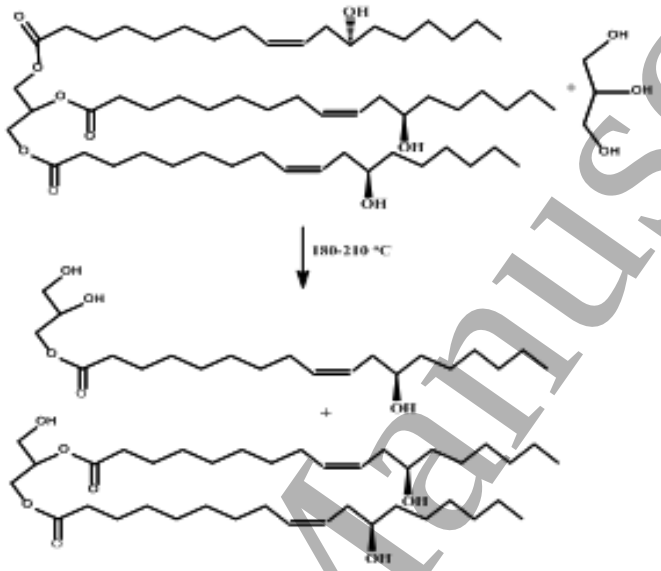


Figure 1 : A schematic representation of the castor oil modification by transesterification process

Table1: Different formulations of foam (pbw)

Samples	Flame Retardants	Mpol	TEDA	n-Pentane	Silicon Oil	MDI
AF-00	Neat	100	2	15	3	110
AF-01	Melamine	100	2	15	3	110
AF-02	TDCPP	100	2	15	3	110
AF-03	TCP	100	2	15	3	110

Accepted Manuscript

3. Results and Discussion

3.1 Density

Density of the resulted RPUFs considerably affect their surface and mechanical properties. The density of the resulted RPUFs samples was determined according to ASTM D 1622-03. The size of the specimen was taken as 30 x 30 x 30 mm^3 and average result of four samples are reported. The density of all samples obtained from different formulations of flame retardants are as given in Table 2. As per the results obtained, the density of the RPUFs incorporated with melamine flame retardant was observed to be decreased as compared to neat RPUFs. The reason for this might be the voids formation in the cell structures of the foam, which, also affected the density of the foam significantly. This result is similar to that already reported in previous literature study [13]. Unlike melamine, the density of RPUFs filled with TDCPP and TCP was found to be more than neat RPUFs which is due to the reduction in the average cell size and the development of uniform cell structure that is also evident from SEM analysis.

Table 2: Density of RPUFs in different formulations

Formulation	Density (kg/m^3)
AF00	102
AF01	82
AF02	112
AF03	107

Accepted Manuscript

3.2 FTIR

A Fourier transform infrared (FTIR) spectrophotometer was employed to analyze the structure of the modified polyols, MDI and polyurethanes. On Nicolet 380 spectrometer, spectra were recorded in the range 400-4000 cm^{-1}. For analysis, powdered samples obtained by crushing the resulted foam were mixed with pure dried KBr to form transparent KBr pellets. Figure 2 (a) shows the characteristic absorption peaks of polyols and MDI at 3402 cm^{-1} (O-H stretching frequency), 2925-2851 cm^{-1} (-CH_2 and -CH_3 stretching frequency), 2268 cm^{-1} (-NCO stretching frequency). Figure 2(b) shows the FTIR of the RPUFs incorporated with anti-flaming agents. It is evident from Figure 1(a) and (b) that there is disappearance of absorption peaks at 2268 cm^{-1} (-NCO stretching frequency) in RPUF, AF01, AF02 and AF03 which confirm that MDI is completely reacted with a polyol to form polyurethane linkage.

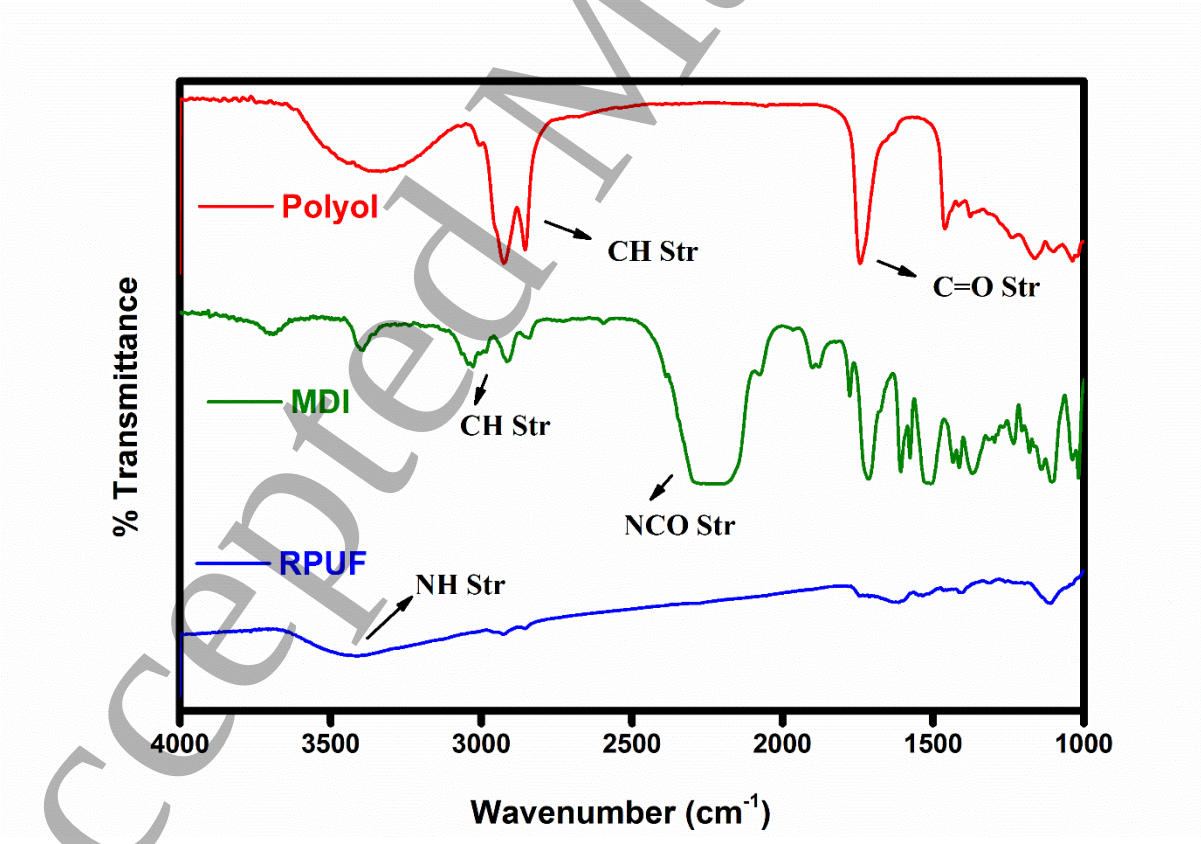


Figure 2 (a): FTIR of Polyol, MDI and Neat RPUF

Accepted Manuscript

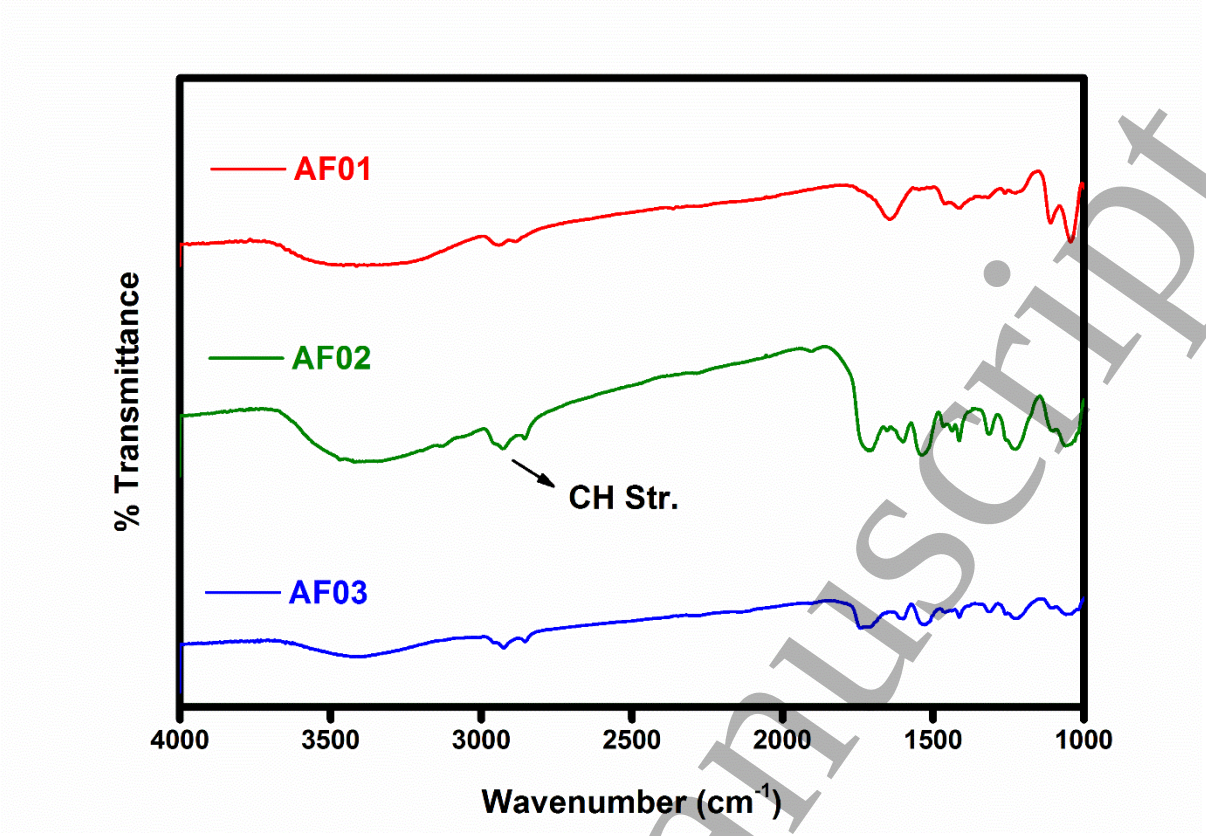


Figure 2 (b): FTIR of formulations AF01, AF02 and AF03

3.3 XRD

XRD patterns for Formulations AF00, AF01, AF02 and AF03 were obtained using BRUKER D8 ADVANCE diffractometer by employing CuKα (Kα –1.54056 Å) radiation. X- ray diffraction pattern of resulted RPUFs incorporated with anti-flaming agents are as shown in Figure 3. From the figure, it is evident that the RPUF exhibit amorphous structure, having broad amorphous peak at 2θ = 20°. Foam samples incorporated with melamine shows slight shift towards crystalline behaviour due to its own crystalline structure.

Accepted Manuscript

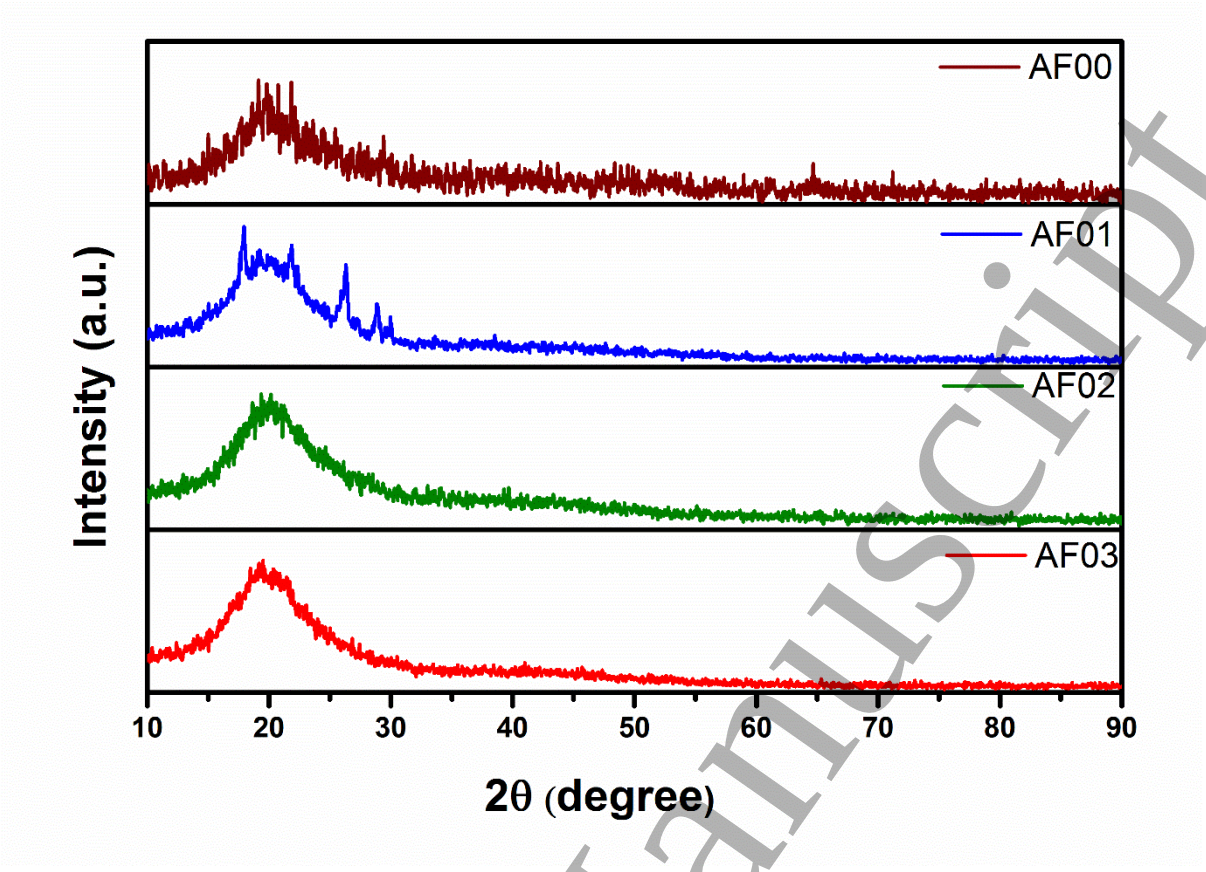


Figure 3: XRD curve of AF00, AF01, AF02 and AF03

3.4 Thermogravimetric Analysis (TGA)

TGA of the foam samples was conducted on Universal V4.5 A, TA instrument using a heating rate of 10 $^{\circ}$C min^{-1} from 30 to 1000 $^{\circ}$C under N_2 atmosphere. Testing of all foam samples was conducted using a gas flow rate of 30 mL/min. Continuous loss in mass was examined as a function of time and temperature. Figure 4 indicates the Thermogravimetric (TG) curves of the foam samples under investigation. The sample AF00 i.e. neat PU (without any flame retardant) clearly exhibit three degradation stages. In the first stage between 80 and 140 $^{\circ}$C, there is a loss of weight because of the evaporation of water from the sample. Second degradation step is visible in the 70 to 240°C temperature range, which is, probably due to the breaking of the urethane linkage. The third degradation step is due to decomposition of material occurring above 250 ^{0}C.

Accepted Manuscript

Compared with the formulation AF00, the other formulations AF01 (Melamine), AF02 (TDCPP) and AF03 (TCP) have a significantly similar decomposing trend. PU foams incorporated with TDCPP shows a higher thermal stability while PU foam with melamine composition shows the least stability.

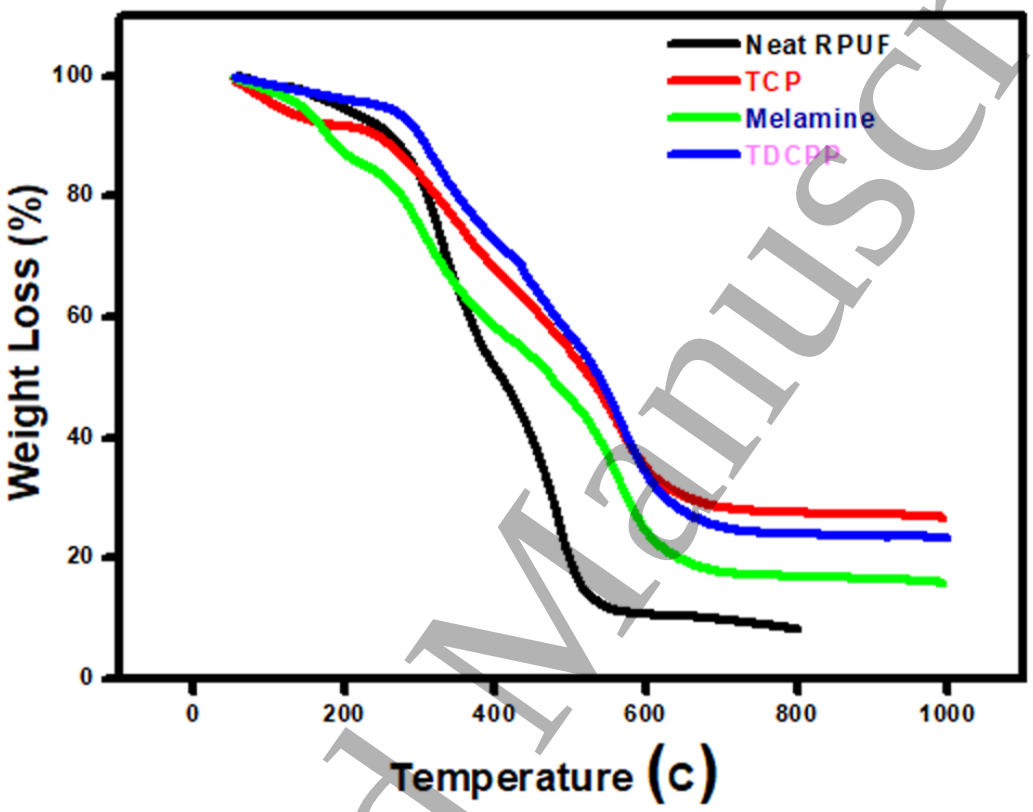


Figure 4: TGA thermograms of the samples AF00, AF01, AF02 and AF03

3.5 Flammability

Flammability of the resulted RPUF samples was classified according to the UL-94 vertical burning standard test [28-30]. The flammability of RPUFs was tested as per the ASTM D 3014 standards. The specimens were mounted vertically on the stand and ignited with the help of a burner for 10 sec. As a standard procedure, a 3/4th inch high blue flame was used for heating at the bottom of the sample for exactly 10 sec and then withdrawn, and again applied for more 10 sec time. The time of glowing and flaming

Accepted Manuscript

was noted as the sample was extinguished. To examine the ignitability of dripping material, a layer of cotton was placed beneath sample, during the test period. A record of time of burning, flame height and weight loss of the samples was accounted and reported in Table 3. From Table 3, it is clear that for the sample with formulation AF00, the recorded flame height travelled is more and for AF02 flame height was found to be least. With AF00 formulation extinguishing time is higher thus shows continuous burning. The reason may be the irregular and improperly disturbed cells in neat PU, with entrapped air, that supports prolong burning. As more is the extinguishing time, more is the recorded weight loss in the samples. Anti-flaming properties of RPUFs were found to be significantly enhanced by the incorporation of flame retardants due to better homogenization of foam structure as per the reported literature [17]. TDCPP incorporated foam samples i.e. formulation AF02 exhibited the best anti-flammable behaviour among the all three formulations. Tris (1,3-dichloro-2-propylphosphate is a semi-volatile flame retardant having both phosphate and chlorine, contributing towards the enhancement of anti-flammability of the foam. TDCPP shows two types of flame retardancy mechanisms i.e. char yield and gas phase from chlorine radical. The phosphate group promotes the char-forming property of the RPUF and the halogenated group i.e. chlorine act in the vapour phase by liberating the chlorine atoms into the flame where they disrupt the combustion reactions [19]. As, the flame retardancy depends upon the content of phosphorus and chlorine in the RPUF, this is the reason that AF02 show the best flame retardant material than AF00, AF01 and AF03.

Accepted Manuscript

Table 3: Flammability Testing of RPUFs with different formulations

Formulation	Flame Height (mm)	Extinguishing Time (sec)	Loss of Weight (grams)	UL – 94 Ratings
AF00	35	Continuously Burn	5.2	No Ratings
AF01	28	82	3.1	V1
AF02	22	55	2.8	V1
AF03	25	62	2.9	V1

3.6 Mechanical Properties

Tensile, compressive and flexural properties of the resulted foams were measured using Instron (Model No.: 3369) universal testing machine as per ASTM D 638, ASTM D 695 and ASTM D 790, respectively. Hardness (Shore D) was determined by using ASTM D 2240 method. The results obtained are as given in Figure 5. As seen in the figure, the neat RPUF (AF00) showed an average value of tensile, compressive and flexural strength and average hardness equal to 0.82 MPa, 6.87 MPa, 3.80 MPa and 45 (Shore D) respectively. Average tensile strength of the RPUF (AF01), (AF02), (AF03) is equal to 0.69 MPa, 2.02 MPa, 1.10 MPa, respectively. The average compressive strength of RPUF (AF01), (AF02), (AF03) was recorded to be 7.24 MPa, 9.12 MPa, 7.42 MPa, respectively. The average flexural strength of RPUF (AF01), (AF02), (AF03) was observed to be 6.30 MPa, 12.02 MPa, 8.42 MPa, respectively and average hardness achieved was 48, 53, 48 (Shore D) respectively. Almost linear increment in the said properties (except tensile strength in case of formulation AF01) up to formulation AF02 indicates a good distribution of antiflamming agents in the PU matrix, leading to

Accepted Manuscript

appropriate adhesion. From AF02 to AF03 formulation, the observed mechanical properties were decreased. The decrease in tensile strength in case of the AF01 formulation is due to involvement of melamine in the polyurethane matrix that weakens the cellular structure. Higher particles size of the melamine results in collapsing of cellular structure during the formation of foam [31]. Significant enhancement in mechanical properties of final product is observed as a result of homogenized dispersal of TDCPP into the polyol. Uniform cell morphology and small pore size of resulted RPUFs with formulation AF02 further support the enhancement in mechanical properties.

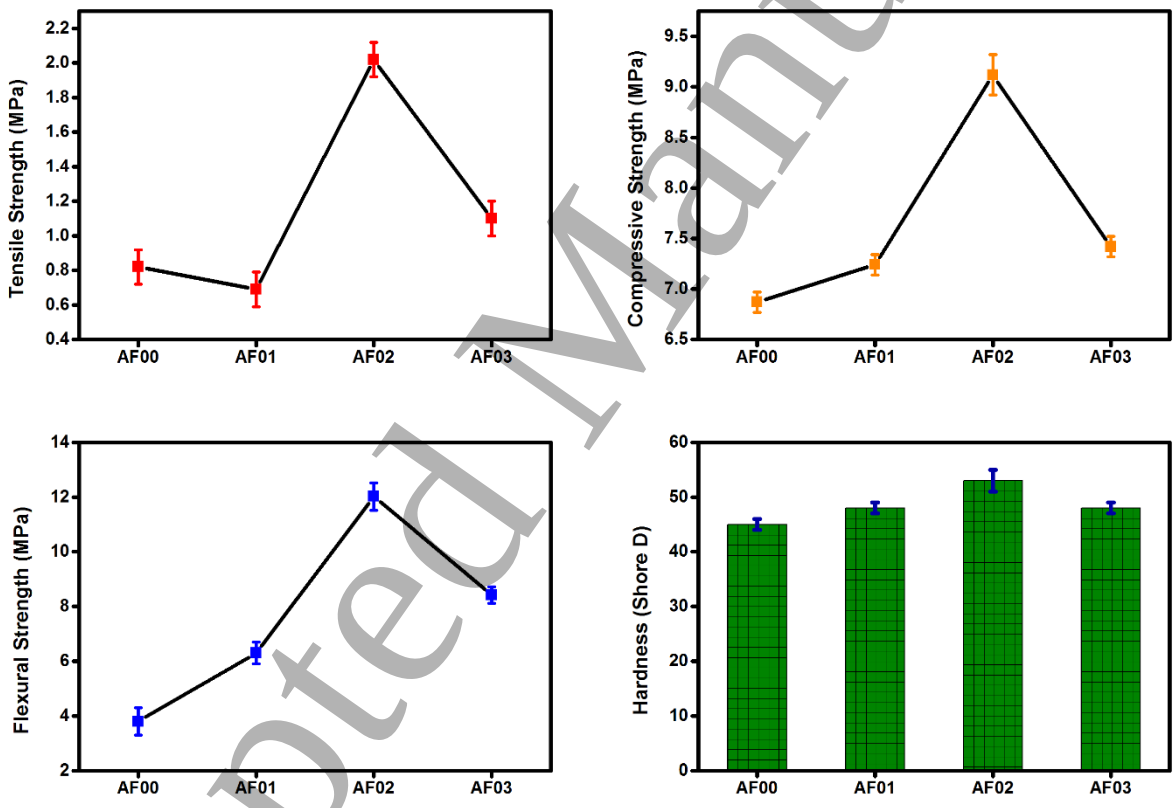


Figure 5: Mechanical properties of different formulations of RPUFs

3.7 SEM Micrograph

The morphologies of RPUFs were observed using scanning electron microscopy (SEM, Hitachi S3700) with an accelerating voltage of 15kV. Prior to testing, samples were gold

Accepted Manuscript

coated to make them conducting. Figure 6 shows the micrographs of the castor oil based rigid polyurethane foams with different formulations. As observed from Figure 6 that the shape of the cells is almost spherical. For the formulations AF00, AF01, AF02 and AF03, the average pore size is equal to 305.25 µm, 266 µm, 225.71 µm and 263.2 µm, respectively. There is a significant reduction in pore size from AF00 to AF02 and then size increase occurs in formulation AF03. The reduced pore size for the formulation AF02 indicated that the 3D structure of the foam was more packed than AF00, AF01 and AF03. The further increase in pore size AF03 is may be due to collapsing of cells. A graph of cell size vs different formulations of RPUFs is as shown in Figure 7. The best results were obtained for mechanical properties for AF02 formulation i.e. RPUF incorporated with Tris (1,3-dichloro-2-propylphosphate) (TDCPP).

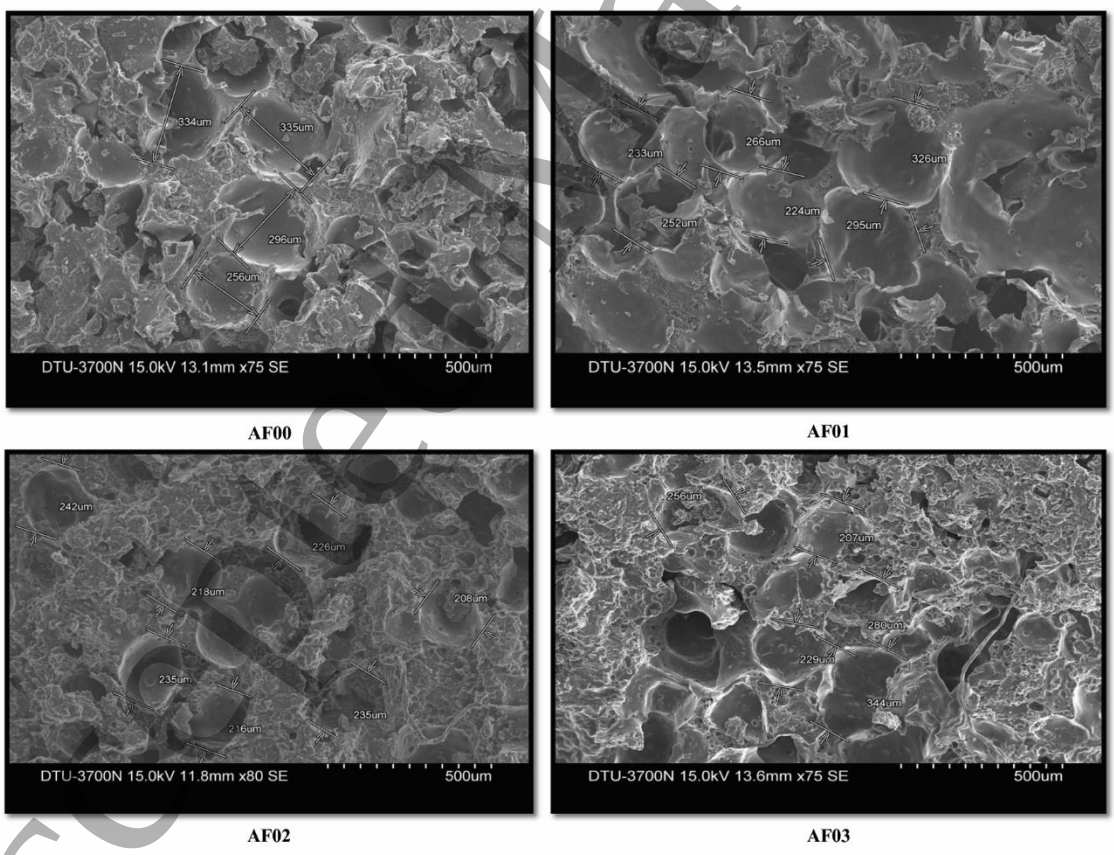


Figure 6: SEM of RPUFs with formulation AF00, AF01, AF02 and AF03

Accepted Manuscript

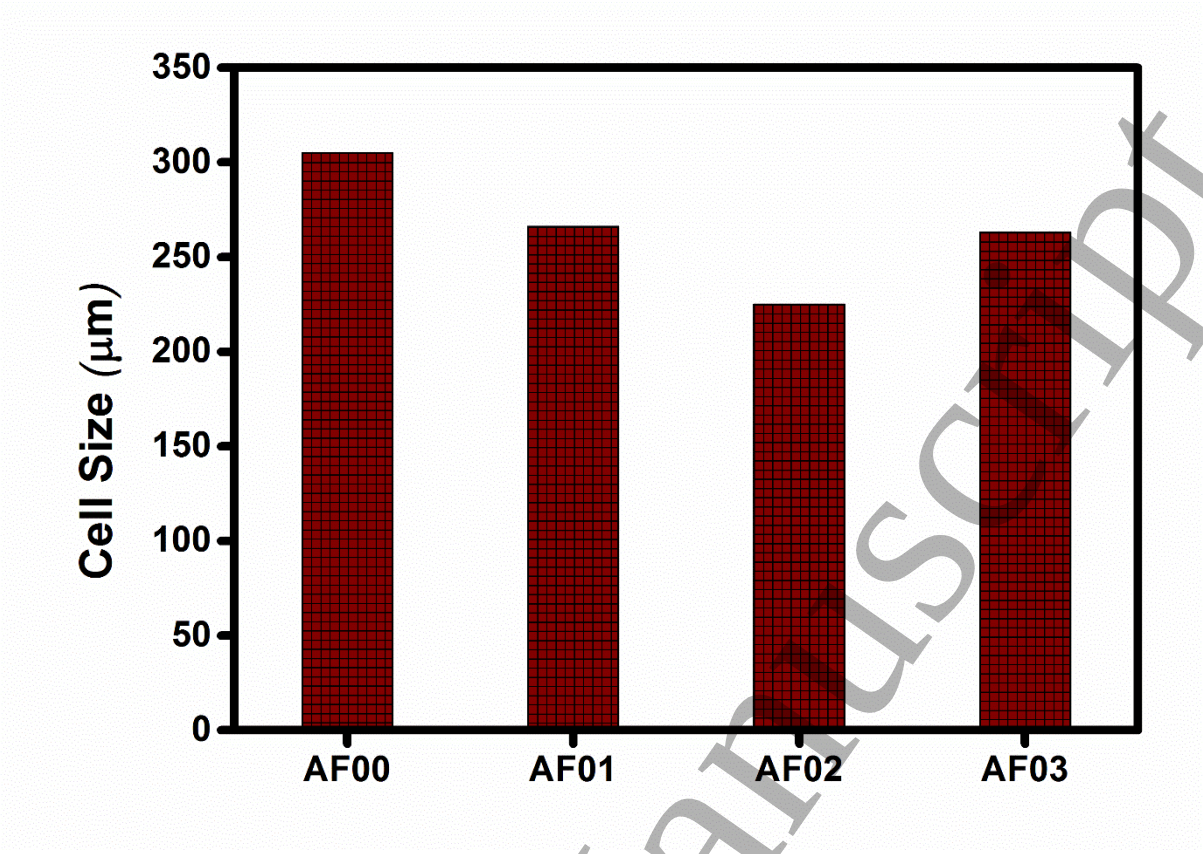


Figure 7: Cell Size of RPUFs with formulation AF00, AF01, AF02 and AF03

4. Conclusion

Though RPUF being one of the industrially important material construction material, it suffers from the drawbacks of its high flammability, formation of dark smoke with the irritating smell and with release of toxic gases on burning. Significant enhancement in the thermal and anti-flamming properties of RPUFs was observed with the addition of anti-flaming agents such as Melamine (2,4,6-Triamino-1,3,5-triazine), Tris(1,3-dichloro-2-propylphosphate) (TDCPP) and Tris (2-chloroethyl) phosphate (TCP).. RPUF incorporated with TDCPP was found to have exhibited best anti-flaming behaviour due to well homogenization of flame retardant into the polyol. Tensile, compressive & flexural strength and Hardness (Shore D) of RPUFs are found to be improved with TDCPP addition because of uniform cell structure and decreased in pore

Accepted Manuscript

size of resulted RPUFs. Tensile strength was found to be poor with melamine based RPUFs as embedment of melamine in the polyurethane matrix weakens the structure due to the formation of inconsistent structure. The resulted foams incorporated with different antiflamming agents with their improved mechanical, thermal and anti-flammable properties are expected to be much suitable for the new generation anti-flammable lightweight materials for different engineering and industrial applications especially in the field of construction and building materials.

Acknowledgement: A sincere thanks to the administration Delhi Technological University for providing facilities to carry out this study. Thanks to Dr. Tyagi, Shivathene Linopack Ltd. Parwanoo, Himachal Pardesh for providing chemicals.

Conflict of Interest: This study has been conducted for the research purpose only and there is no conflict of interest is involved.

References:

1) R. Yang, W. Hu, L. Xu, Y. Song, J. Li, Synthesis, mechanical properties and fire behaviours of rigid polyurethane foam with a reactive flame retardant containing phosphazene and phosphate, Polym. Degrad. Stab. 122 (2015) 102-109. doi:10.1016/j.polymdegradstab.2015.10.007.
2) H. Hatakeyama, H. Matsumura, T. Hatakeyama, Glass transition and thermal degradation of rigid polyurethane foams derived from castor oil-molasses polyols, J. Therm. Anal. Calorim. 111 (2013) 1545-1552. doi:10.1007/s10973-012-2501-5.
3) F. Luo, K. Wu, M. Lu, S. Nie. X. Guan, Thermal degradation and flame retardancy of microencapsulated ammonium polyphosphate in rigid polyurethane foam, J. Therm. Anal. Calorim. 120 (2015) 1327-1335. doi:10.1007/s10973-015-4425-3.
4) S.V. Levchik, E.D. Weil, Thermal decomposition, combustion and fire-retardancy of polyurethanes – A review of the recent literature, Polym. Int. 53 (2004) 1585-1610. doi:10.1002/pi.1314.

Accepted Manuscript

5) U. Almeida Pinto, L.L.Y. Visconte, J. Gallo, R.C.R. Nunes, Flame retardancy in thermoplastic polyurethane elastomers (TPU) with mica and aluminium trihydrate (ATH), Polym. Degrad. Stab. 69 (2000) 257-260. doi:10.1016/S0141-3910(00)00047-1.

6) H.Guo, Q.Gao, C.Ouyang, K. Zheng, W. Xu, Research on properties of rigid polyurethane foam with heteroaromatic and brominated benzyl polyols, J. Appl. Polym, Sci. 132 (2015) 1-8. doi:10.1002/app.42349.

7) W. Zatorski, Z.K. Brzozowski, A. Kolbrecki, New developments in chemical modification of fire-safe rigid polyurethane foams, Polym. Degrad. Stab. 93 (2008) 2071-2076. doi:10.1016/j.polymdegradstab.2008.05.032.

8) H. Zhu, Z. Peng, Y. Chen, G. Li, L. Wang, Y. Tang, R. Pang, Z.U.H. Khan, P. Wan, Preparation and characterization of flame retardant polyurethane foams containing phosphorus-nitrogen-functionalized lignin, RSC Adv. 4 (2014) 55271-55279. doi:10.1039/C4RA08429B.

9) L. Qian, F. Feng, S. Tang, Bi-phase flame-retardant effect of hexa-phenoxy-cyclotriphosphazene on rigid polyurethane foams containing expandable graphite, Polym. (United Kingdom). 55 (2014) 95-101. doi:10.1016/j.polymer.2013.12.015.

10) L. Ye, X.Y. Meng, X. Ji, Z.M. Li, J.H. Tang, Synthesis and characterization of expandable graphite-poly(methyl methacrylate) composite particles and their application to flame retardation of rigid polyurethane foams, Polym. Degrad. Stab. 94 (2009) 971-979. doi:10.1016/j.polymdegradstab.2009.03.016.

11) A. Agrawal, R. Kaur, R. S. Walia, Flame Retardancy of Ceramic Based Rigid Polyurethane Foam Composites, Jl App. Poly. Sci., 2019, DOI: 10.1002/app.48250.

12) A. Agrawal, R. Kaur, R. S. Walia, Investigation on Flammability of Rigid Polyurethane Foam-Mineral Fillers Composite, Fire and Materials, 2019, DOI: 10.1002/fam.2751.

13) M. Thirumal, D. Khastgir, G.B. Nando, Y.P. Naik, N.K. Singha, Halogen-free flame retardant PUF: Effect of melamine compounds on mechanical, thermal and flame retardant properties, Polym. Degrad. Stab. 95 (2010) 1138-1145. doi:10.1016/j.polymdegradstab.2010.01.035.

14) S.Y. Lu, I. Hamerton, Recent developments in the chemistry of halogen-free flame retardant polymers, Prog. Polym. Sci. 27 (2002) 1661-1712. doi:10.1016/S0079-6700(02)00018-7.

15) H. Horacek, R. Grabner, Advantages of flame retardants based on nitrogen compounds, Polym. Degrad. Stab. 54 (1996) 205-215. doi:10.1016/S0141-3910(96)00045-6.

Accepted Manuscript

16) K. Wu, L. Song, Z. Wang, Y. Hu, Preparation and characterization of double shell microencapsulated ammonium polyphosphate and its flame retardance in polypropylene, J. Polym. Res. 16 (2009) 283-294. doi:10.1007/s10965-008-9228-9.

17) J. Czech-polak, M. Heneczkowski, B. Przybyszewski, A. Czulak, M. Gude, Effect of environmentally-friendly flame retardants on fire resistance and mechanical properties of rigid polyurethane foams, Polyimery/Polymers 61 (2016) 113-116.

18) R.V. Rao, B.J. Manujesh, Burning behaviour of polyurethane foam cored/E-glass reinforced/vinylester sandwich composites: Effect of core density and fibre architecture, European Journal of Applied Engineering and Scientific Research 2 (2013) 1-8.

19) E.S. Goda, K.R. Yoon, S. H. El-Sayed, S.E. Hong, Halloysite nanotubes as smart flame retardant and economic reinforcing materials: A review, Thermochimica Acta 669 (2018), 173-184.

20) M. Malik, R. Kaur, Synthesis of NIPU by the carbonation of canola oil using highly efficient 5,10,15-tris(pentafluorophenyl) corrolato-manganese(III) complex as novel catalyst, Polym. Adv. Tech., 2018, 29(3),1078-1085, DOI. 10.1002/pat.4219

21) A. Aggarwal, R.,Kaur, R.S.Walia, PU Foam Derived from Renewable Sources: Perspective on Properties Enhancement: An Overview, Europ. Polym. Jl,2017 95, 255-274, DOI.10.1016/j.eurpolymj.2017.08.022.

22) M. Kumar, R. Kaur, Glass fiber reinforced rigid polyurethane foam: synthesis and characterization, e-Polymers, 2017,7(6),517-521, DOI. 10.1515/epoly-2017-0072.

23) M. Malik, R. Kaur, Influence of aliphatic and aromatic isocyanates on the properties of polyether-ester polyol based PU adhesive system, Polym. Eng. Sci., 2018, 58 (1), 112-117, DOI.10.1002/pen.24537.

24) M. Malik, R. Kaur, Mechanical and Thermal Properties of Castor Oil–Based Polyurethane Adhesive: Effect of TiO_2 Filler, Adv. Polym. Tech., 2018, 37(1),24-30, DOI. 10.1002/adv.21637.

25) R. Kaur, M. Kumar, Function of Silicon Oil in the Castor Oil Based Rigid Polyurethane Foams", Jl. Polym. Eng., 2013, 33 (9), 857-861, doi:10.1515/polyeng-2013-0083.

26) M. Kumar, R. Kaur, Effect of Different Formulations of MDI on Rigid Polyurethane Foams based on Castor Oil, Int. Jl. Sci. Res. Review., 2013, 2 (1) Special Issue April – May, 29-42.

Accepted Manuscript

27) A. Agrawal, R. Kaur, R. S. Walia, Vegetable Oil Based Conducting Rigid PU Foam, e-Poly., 2019, 19, 411-420. DOI: https://doi.org/10.1515/epoly-2019-0042.

28) T. Tirri, M. Aubert, C.E. Wilen, R. Pfaendner, H. Hoppe, Novel tetrapotassium azo diphophonate (INAZO) as flame retardant for polyurethane adhesives, Polym. Degrad. Stab. 97 (2012) 375-382. doi:10.1016/j.polymdegradstab. 2011.12.005.

29) Y. Wang, J. Zhang, Thermal stabilities of drops of burning thermoplastics under the UL 94 vertical test conditions, J. Hazard. Matter. 246-247 (2013) 103-109. doi:10.1016/j.jhazmat.2012.12.020.

30) H. Wu, M. Lee, Y. Wu, Y. Su, C.M. Ma, Pultruded fiber-reinforced polyurethane-toughned phenolic resin.II. Mechanical properties, thermal properties, and flame resistance, J. Appl. Polym. Sci. 62 (1996) 227-234. Doi:10.1002/(SICI)1097-4628(19961003)62:1<227::AID-APP26>3.0.CO;2-0.

31) A. Kong, U. Fehrenbacher, T. Hirth, E. Kroke, Flexible Polyurethane Foam with the Flame-retardant Melamine, J. Cell. Plast. 44 (2008) 469-480. Doi:10.1177/0021955X08095502.

Figure Caption:

Figure 1: A schematic representation of the castor oil modification by transesterification process

Figure 2 (a): FTIR of Polyol, MDI and Neat RPUF

Figure 2 (b): FTIR of formulations AF01, AF02 and AF03

Figure 3: XRD curve of AF00, AF01, AF02 and AF03

Figure 4: TGA thermograms of the samples AF00, AF01, AF02 and AF03

Figure 5: Mechanical properties of different formulations of RPUFs

Figure 6: SEM of RPUFs with formulation AF00, AF01, AF02 and AF03

Figure 7: Cell Size of RPUFs with formulation AF00, AF01, AF02 and AF03

Accepted Manuscript

Materials Research Express

ACCEPTED MANUSCRIPT

Analysis of mechanical behaviour of L-arginine Hydrobromide Monohydrate (LAHBr) single crystal grown by unidirectional growth technique

To cite this article before publication: Kanika Thukral *et al* 2020 *Mater. Res. Express* in press https://doi.org/10.1088/2053-1591/ab6b5b

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is **© 2020 IOP Publishing Ltd**.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.
As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by-nc-nd/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 14.139.251.100 on 20/01/2020 at 04:23

Analysis of mechanical behaviour of L-arginine Hydrobromide Monohydrate (LAHBr) single crystal grown by unidirectional growth technique

Kanika Thukral[1, 2,3 $], Sonia[1, 2,4, $], Neema Jha[5], N. Vijayan[6*], Budhendra Singh[6], S. A. Martin Britto Dhas[7]

[1]Academy of Scientific and Innovative Research, CSIR- National Physical Laboratory New Delhi-110012, India

[2]CSIR-National Physical Laboratory, Dr. K.S. Krishnan Road, New Delhi–110012, India

[3]Department of Applied Physics, Delhi Technological University, Bawana Road, New Delhi 110 042, India

[4]Department of Sciences, Baba Mastnath University, Asthal Bohar Rohtak-124001, India

[5]Guru Gobind Singh Indraprastha University, Sector-16, Dwarka, New Delhi-110078, India

[6]TEMA-NRD, Mechanical Engineering Department and Aveiro Institute of Nanotechnology (AIN), University of Aveiro, 3810-193 Aveiro, Portugal

[7]Department of Physics, Sacred Heart College, Tirupattur, Vellore-635601, India

Abstract

A single crystal of L-arginine Hydrobromide Monohydrate (LAHBr) with excellent crystalline perfection was grown by a modified way of unidirectional technique. The criteria behind this modified version were to reduce the temperature gradient of the solution which will enhance the quality of the single crystal. The formation of the crystal was confirmed by calculating lattice parameters from powder X-ray diffraction technique. The quality of the single crystal was analyzed by high-resolution X-ray diffraction which suggests that single crystal was free from grain boundaries and have the value full width half maxima is 9.36 arc second. Owing to high crystallinity, the single crystal was further underwent mechanical analysis using nanoindentation technique. Its structural, as well as physical transformation have been assessed by laser and shock damage threshold techniques. The presence of surface defects was examined by scanning electron microscope (SEM) and it is in tune with the high resolution X-ray diffraction method.

Keywords: X-ray diffraction, Crystal growth, Laser, Mechanical stability, Nanoindentation

Accepted Manuscript

[$] **Both authors have equal authorship or contribution in research work**

[*]**Corresponding Author**

Dr. N. Vijayan, Principal Scientist

Indian Reference Materials Division

CSIR-National Physical Laboratory

Dr. K.S. Krishnan Road, New Delhi-110012, India

E-mail: nvijayan@nplindia.org; vjnphy@yahoo.com

Ph. No. +91-11-45608263; FAX No. +91-11-45609310

Introduction

It is the demand of the time when our technology has switched to photons from electrons for the transmission of information. In this view, many new materials with unique optical properties have been analyzed to fulfill the current day to day requirements [1]. In search of innovative materials, a new class has been found i.e. semi-organic materials in which a polarizable organic compound made abound with the inorganic host material. Such type of materials is mechanically more stable than organic materials and also having high laser damage threshold value [2]. The semi-organic materials can be easily grown in three dimensions by solution growth technique which further can be cut and polished for application usage. But the main drawback associated with this technique was its inefficiency in growing bulk size crystal. Also, there has been a high probability of spontaneous nucleation during the growth period which ruins the quality of single crystal [3]. Using conventional growth method, it was found that the single crystals of different orientation are grown along with different morphology. Before making any fabrication of devices, it is necessary to obtain a good quality and bulk size single crystal along a particular direction [4-5]. In this direction, a technique has been discovered to grow unidirectional crystal from a solution called Sankaranarayanan- Ramasamy (SR) method. This method has 100% efficiency to convert entire amount of solute into crystal and thus, avoide the wastage of solution. The direction of grown crystal was decided by suitably mounting the seed in the neck of ampoule.

L-arginine is an amino acid which has the tendency to crystallize in non-centrosymmetric structure and have anisotropic nature. It means that the crystal of these compounds shows

Accepted Manuscript

different behavior in different directions. So, it is worthwhile if the crystal will grow through this technique. Because through this technique it is possible to grow crystal in a direction that shows high efficiency. There are many compounds of L-arginine that have been reported who show non-linear optical property like L-arginine phosphate (LAP), L-arginine trifluoroacetate (LATF), L-argininium bis(trifluoroacetate) (LABTF), L-argininium phosphite (LAPI), L-arginine 4-nitrophenolate 4-nitrophenol dihydrate (LAPP), and many more [6-10]. Such properties were motivated the researchers to grow and characterize the new compound of this family of crystals. In addition to that, L-arginine family materials are having good second harmonic efficiency than that of standard potassium dihydrogen phosphate (KDP) [11] single crystals.

In the present work, LAHBr single crystal having dimension 4.5 cm lengh and 1.5 cm diameter has been grown by using SR method in (001) direction. The present report mainly deal with the mechanical stability of LAHBr compound. The grown ingot was subjected to high resolution X-ray diffraction which revelas the presence of grain bouandaries present in the the crystal. Further, Photoluminscence technique is also employed on LAHBr single crystal to identify the crystal defects. Whereas, nanoindentation provides the mechanical strength of the grown crystal, against different loads. Laser damage threshold of the title compound was examinded by using Nd:YAG laser as the source. The detailed discussions were given in the forthcoimg sections.

Experimental method

Apart from the conventional method, a new design of SR method has been used to grow crystal in order to improve the crystalline quality of the sample. Initially, L-arginine Hydrobromide Monohydrate (LAHBr) was synthesized by taking stoichiometric ratio of commercially available L-Arginine (molecular weight 174.20 and melting point 200˚C) and hydrobromic acid (molecular weight 80.91) and then dissolved in a double-distilled water. The chemical scheme for the present sysnthesis is given below:

$$C_6H_{14}N_4O_2 + HBr + H_2O \rightarrow C_6H_{15}N_4O_2Br + H_2O$$

Further, a seed crystal has been developed by a slow evaporation method and then a V-type shape has been given to the seed so that it can fit into the neck of the ampoule. After that, ampoule was kept in a glass container filled with water and set at a temperature of 23 deg. C. The heating coils were set inside the tub to provide the proper temperature. Further, the

Accepted Manuscript

temperature was controlled and continuously monitored by temperature controller along with the help of a temperature sensor. In the top part of the ampoule, a cup-like design has been made. The reason for such design was to avoid multi-nucleation as there was some temperature difference between the top and bottom of the ampoule. In order to reduce that temperature difference, the cup was filled with water and T shaped heater has been induced in it for maintaining the temperature. By reducing the temperature difference sudden formation of crystallites have been avoided. Fig. 1 shows the schematic diagram of the setup.

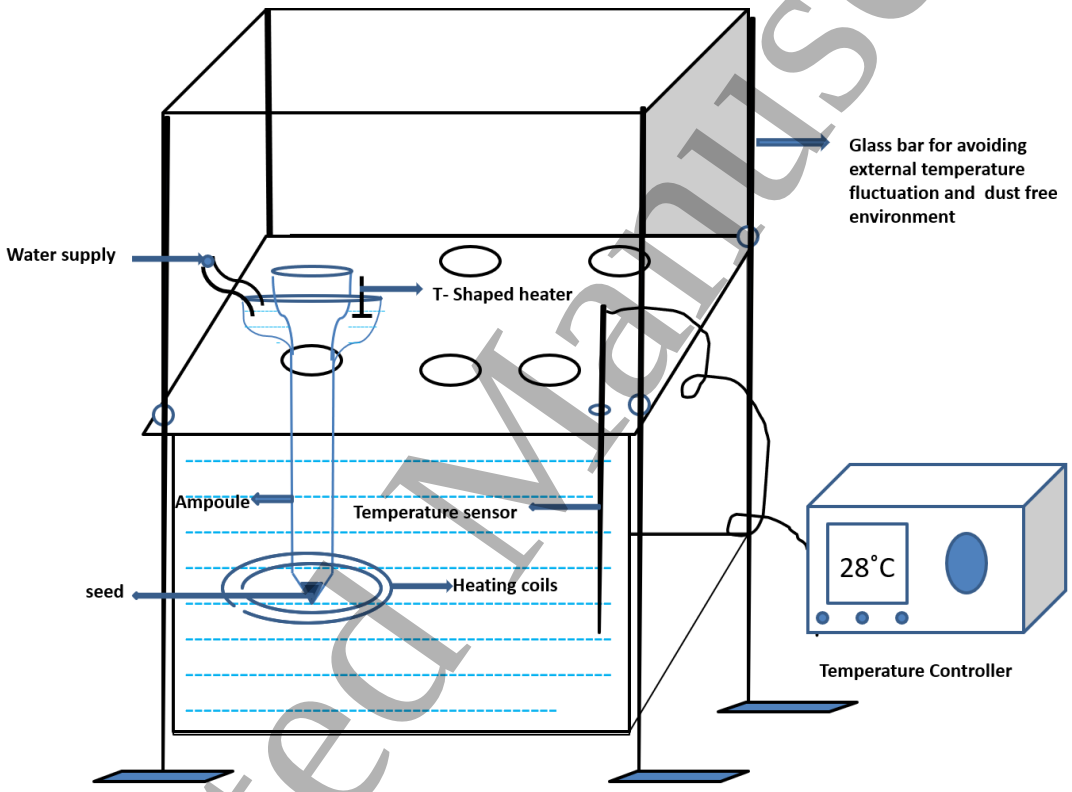


Fig. 1 Schematic diagram for SR Method

After a span of one month, a unidirectional bulk-size single crystal was harvested which is shown in Fig. 2. Further, the crystal was cut and polished into a number of smaller crystals for characterization purposes.

Accepted Manuscript

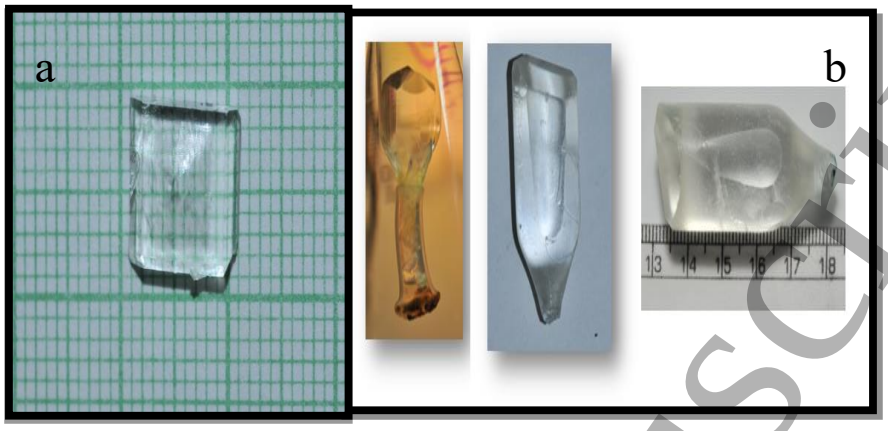


Fig. 2(a) Seed crystal; **(b)** unidirectional grown single crystal

Characterization techniques

Powder X-ray diffraction (PXRD):-

The primary investigation for the formation of the compound can be done by obtaining powder X-ray diffraction pattern. This process includes phase identification of the crystalline material which reveals the unit cell parameter of the material. The diffraction pattern can be attained by interference of the radiation, passes through the compound carries structural information. X-rays are the most efficient tool for compound identification as its wavelength is comparable to the lattice parameters of the unit cell. An excellent quality crystal was crushed in the form of fine powder which then subjected to Rigaku X-ray diffractometer using CuK_{α} as a source to produce incident radiation (λ~1.54Å) with a scan speed of 4º/min. The crystallites are arbitrarily oriented with reference to the X-ray beam. Therefore, maximum reflections can be obtained from all set of lattice planes which satisfies Bragg's condition i.e. $n\lambda=2d\sin\theta$, where λ is incident wavelength, d was spacing between planes and θ was the diffraction angle. The powdered sample has been scanned over a range of 2θ i.e. 10-40°, to attain diffraction planes. The observed pattern is shown in Fig. 3. The calculated lattice parameters is a = 11.2042 Å, b = 8.5758 Å, c = 11.2210 Å, α=ϒ=90˚, β=91.120˚, V=1077.95 $Å^3$ which is good in agreement with the reported literature [12,13].

Accepted Manuscript

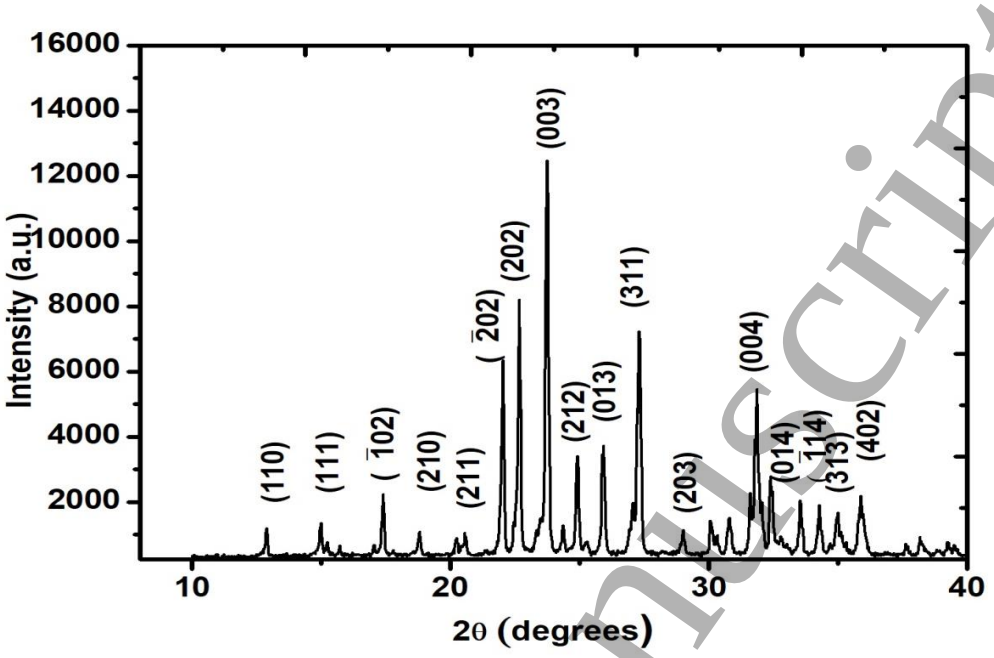


Fig. 3 Powder X-ray diffraction pattern

High- resolution X-ray diffraction (HRXRD):-

High- resolution X-ray diffraction is an ultimate tool to assess the crystalline perfection of the crystal. A PANalytical X'Pert PRO MRD high-resolution XRD system has been used for the scanning of the crystal with $CuK\alpha_1$ radiation. The rocking curves of the crystals for the diffraction planes were recorded in symmetrical Bragg geometry using the natural facets by performing the ω scan with double-axis geometry. The monochromated X-ray beam ($CuK_{\alpha 1}$) falling on the sample was achieved using a four-bounce monochromator Ge (220). Using a scintillation detector, the diffracted beam was detected to estimate the crystallinity of a single crystal along a particular direction which includes structural grain boundaries, dislocations, and other defects. Fig. 4 represents the rocking curve of L-arginine HBr single crystal along the diffracted plane of (001).

Accepted Manuscript

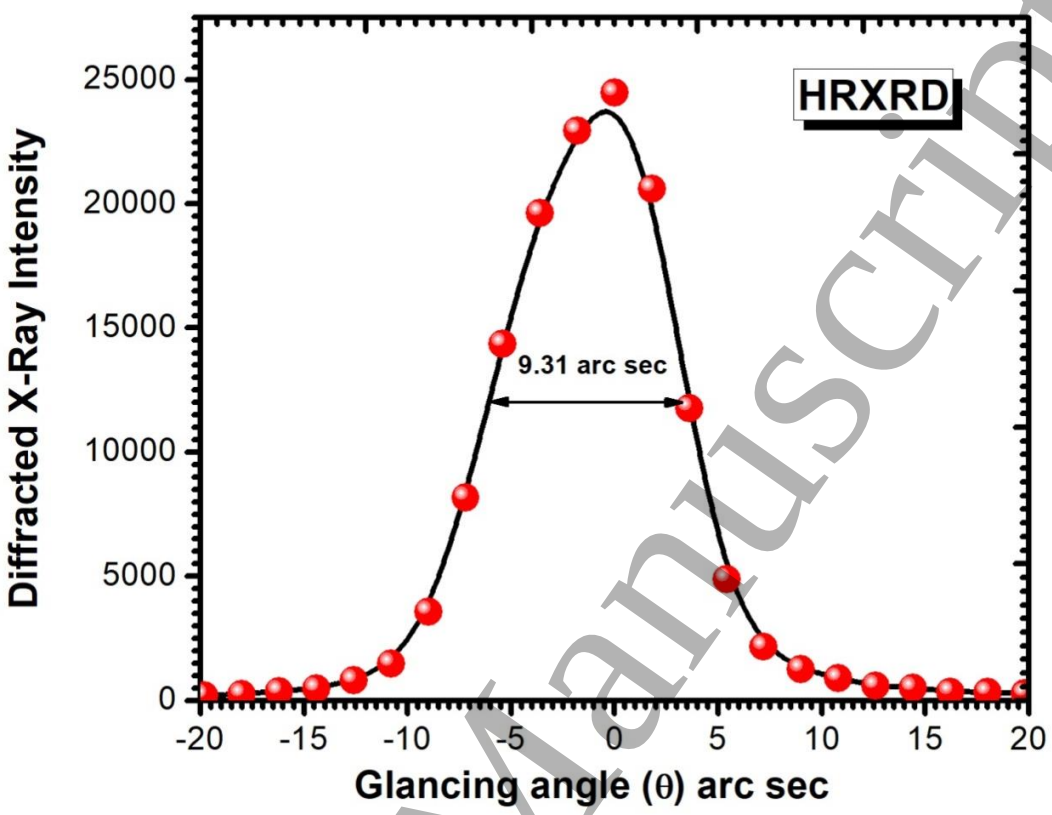


Fig. 4 Rocking curve recorded for a plane (001) of LAHBr single crystal

As the figure shows that the diffracted curve (DC) contains only a single peak which suggests that the crystal is free from grain boundaries. The FWHM of the curve was found to be nearly 9.31 arc second. This value is a little higher than the predicted value for ideal crystals [14-15]. This suggests that the defect concentration is quite low in the crystal which makes it a suitable candidate for their implementation in photonic applications since no sufficient degradation of properties is caused by these defects [16].

Nanoindentation:-

Applications of a non-linear optical material require interactions of materials with lasers. Laser being high energy radiations, generally tends to damage the crystal surface. The extent of these damages depends on the mechanical strength of the material. Therefore, mechanical parameters play a vital role in determining the suitability of material in gadget manufacturing. Nanoindentation is an efficient tool to extract strength parameters and detailed analysis of the

Accepted Manuscript

material's mechanical strength at nanoscale. So, this technique was utilized to measure its hardness, stiffness, young's modulus from the load-displacement curve. In this technique, the sample was subjected to different loads using sharp Berkovich tip (three-sided pyramidal) of edge radius of 20 nm, the indenter faces angle was fixed at 65.3° from the vertical axes and connected to fully calibrated nanoindenter. The High-resolution Atomic Force Microscope (AFM) in contact mode is used to visualize the surface defects or cracks when the load is applied. Using, Oliver and Pharr's method the results were analyzed [17]. Using the recorded data, first we have plotted the load Vs. displacement (loading and unloading) curve using the relation

$$F = \alpha(h - h_f)^m \quad (1)$$

Where F and h_f referred to the applied force on the exposed material and final displacement where the unloading curve completes respectively. α and m are the constants. The slope of the unloading curves give Stiffness can be calculated as

$$s = \frac{dF}{dh} = \alpha m(h - h_c)^{m-1} \quad (2)$$

Using the above value, one can deduce contact depth which can further be used to evaluate the independent hardness of the titled material using equations given below

$$h_c = h_{max} - \varepsilon \times \frac{F_{max}}{S} \quad (3)$$

$$H = \frac{F_{max}}{A_P} \quad (4)$$

Where H, F_{max}, and A_P stand for the load-dependent hardness, maximum load offered by indenter and projected contact area of indenter respectively. The load-independent hardness (H_o) is determined using empirical relation.

$$H_0 = Ka_2 \quad (5)$$

Where K is a constant and its value vary with indenter geometry. Further, Young's modulus can be calculated using the equation

Accepted Manuscript

$$E = \frac{1\sqrt{\pi}}{2A_P}\frac{dF}{dh} \quad (6)$$

In the current study, a good quality single crystal of LAHBr was first lapped and polished to get a smooth surface. After that, the sample was subjected to nanoindentation over a load range of 5 mN to 150 mN. In this measurement, force and displacement are recorded whenever the test material is compressed by the Berkovich tip with a prescribed loading and unloading profile. Fig. 5 shows a P-h curve obtained from (001) plane and the inset shows a pileup from the crystal impression made by 50 mN force.

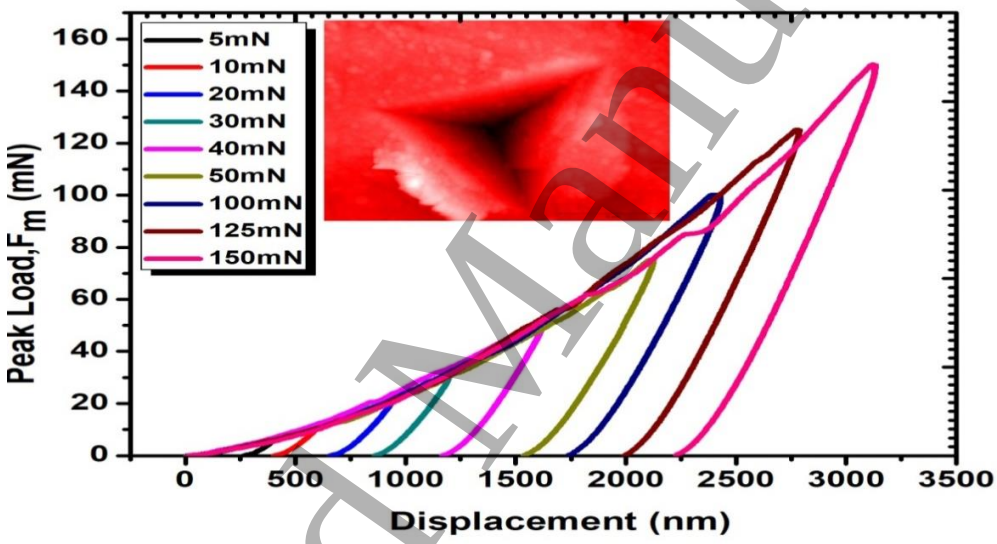


Fig. 5 Load displacement curve obtained from the (001) plane of the LAHBr single crystal with inset of imprint made by 50 mN load

On a detailed analysis of the curve at each load, traces of slight pop-in feature have been observed during the loading sequence. These features are corresponding to nucleation of plastic deformation during the indentation of a single crystal. In addition, no pop-out has been observed in the unloading region which is generally attributed to pressure-induced phase transformation [18]. After applying various loads, the optical image of the test material was captured and compared with a pristine surface as shown in Fig. 6.

Accepted Manuscript

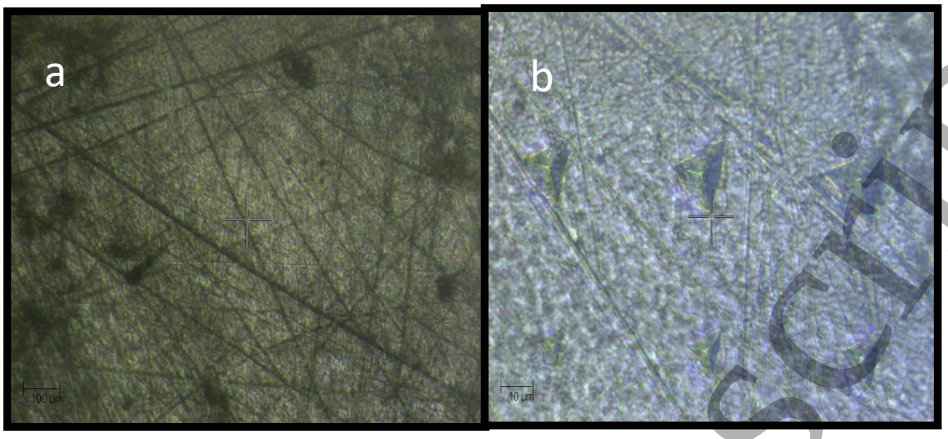


Fig. 6 Optical images of the LAHBr single crystal (a) before and; (b) after indentation

The different mechanical parameters calculated from the load-displacement curve using the above equations are tabulated in Table1. A Contact depth Vs. peak load curve has been plotted to calculate the load-independent hardness of the single crystal as shown in Fig. 7.

Table 1 Various obtained parameters related to mechanical properties of LAHBr single crystal along the (001) plane

(001) Plane										
F (mN)	**H_{it} (MPa)**	**H_V (Vickers)**	**E_{it} (GPa)**	**h_m (nm)**	**S [mN nm^{-1}]**	**h_c (nm)**	**h_r (nm)**	**h_p (nm)**	**A_P (nm^{-2})**	**m**
5.064	1888.483	174.894	31.459	389.575	0.064	330.41	310.595	267.642	2.68146E6	1.513
10.062	1634.5	151.372	26.94	591.356	0.084	500.62	470.919	411.612	6.15574E6	1.462
20.061	1279.532	118.499	22.624	933.503	0.112	798.954	755.05	669.117	1.56786E7	1.456
30.048	1166.607	108.04	20.652	1196.732	0.132	1024.04	968.705	872.051	2.57572E7	1.419
50.069	1061.831	98.337	18.424	1624.427	0.159	1385.565	1310.298	1183.317	4.71539E7	1.389
75.063	941.425	87.186	15.774	2124.423	0.178	1801.718	1702.493	1556.598	7.97329E7	1.355
100.06	986.516	91.362	15.24	2428.792	0.194	2032.111	1912.896	1760.186	1.01428E8	1.324
125.03	944.294	87.452	14.29	2787.805	0.208	2321.783	2186.632	2017.706	1.32406E8	1.285
150.071	901.619	83.5	13.459	3132.037	0.22	2603.188	2449.252	2259.535	1.66447E8	1.288

Accepted Manuscript

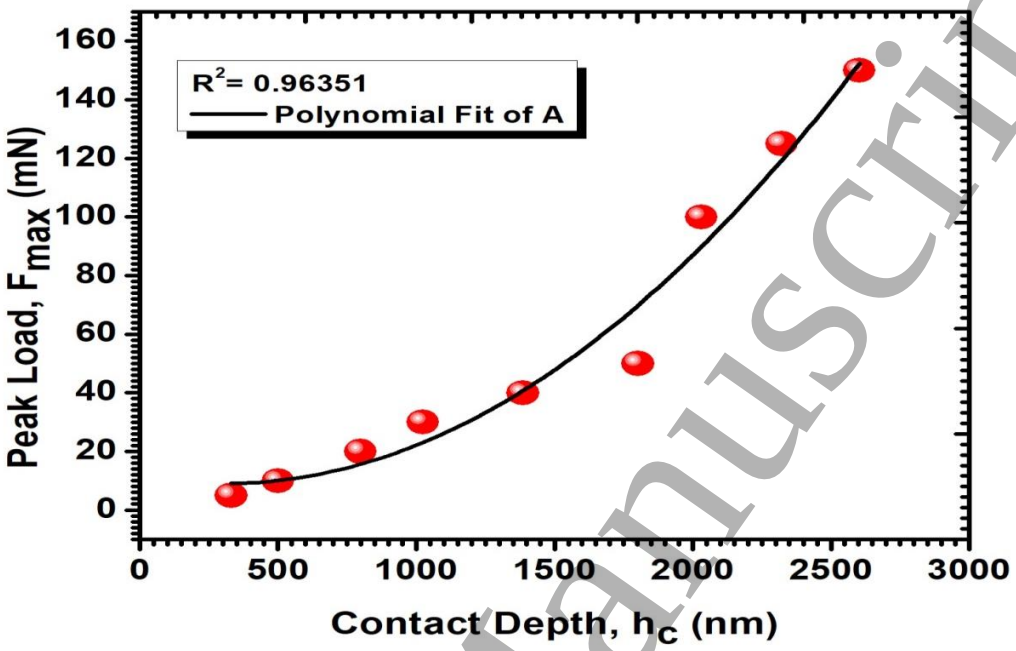


Fig. 7 Curve of the peak load Vs. contact depth for the LAHBr single crystal.

From the curve, it can be clearly seen that there is an increment in the value of contact depth as peak load increases. This is due to the fact that indentation size can affect the hardness of the material. The contact depth, peak load and hardness of the material is related as

$$F_{max} = a_0 + a_1 h_c + a_2 {h_c}^2 \qquad (7)$$

A polynomial data has been fitted to the curve to get the coefficient values a_0, a_1, a_2. The obtained value of a_2 coefficient was 2.71812 $\times 10^{-5}$ mN nm^{-2} and calculated independent hardness is found to be 1.109 GPa or 1109 MPa. This value is found to be quiet high as compared to other compounds of L-arginine such as L-arginine 4-nitrophenolate 4-nitrophenol dihydrate (LAPP), L-argininium bis(trifluoroacetate) (LABTF) [19-20]. Thus the present compound favors the high resistance against the deformation. Further, applying the Oliver and Pharr method, Young's modulus has been extracted from the P-h curve is plotted against peak load as shown in Fig. 8.

Accepted Manuscript

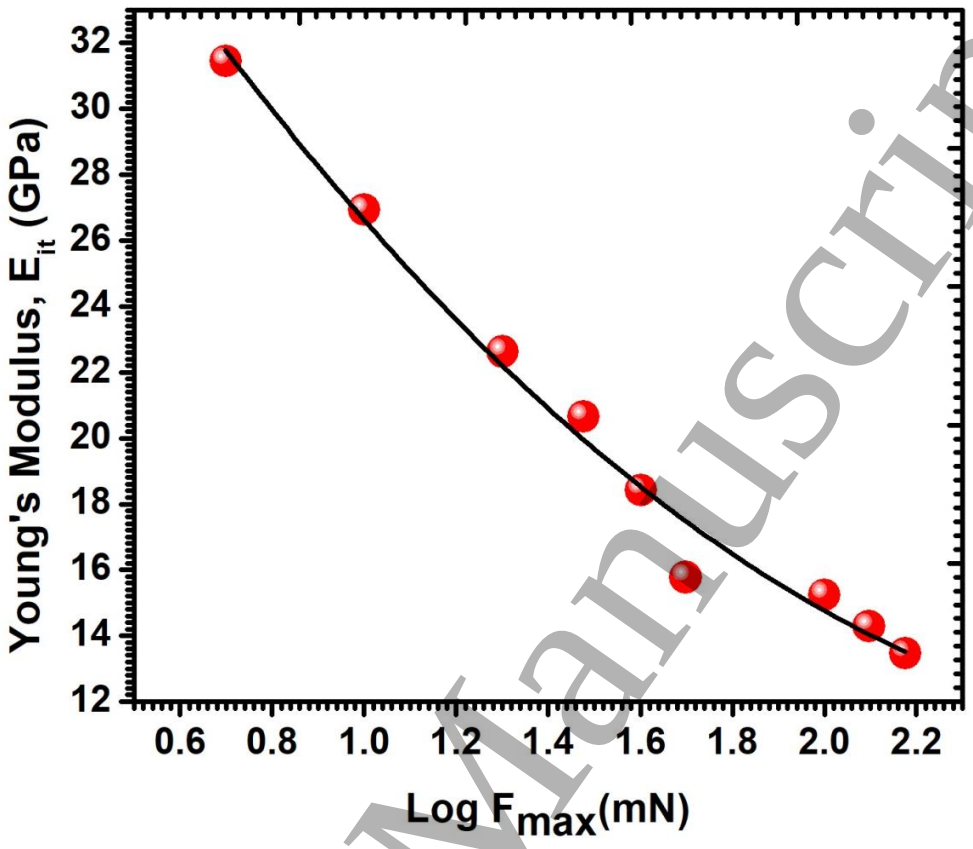


Fig. 8 Variation of Young's modulus as a function of the log of the peak load for the LAHBr single crystal.

The curve indicates that there is a significant decrement in the value of Young's modulus as the peak load increases. The reason for such behavior is due to the induced strain in single crystal while applying the nano loads. Fig. 9 represents the plot of initial unloading stiffness (S) with the contact depth at peak load for the titled compound.

Accepted Manuscript

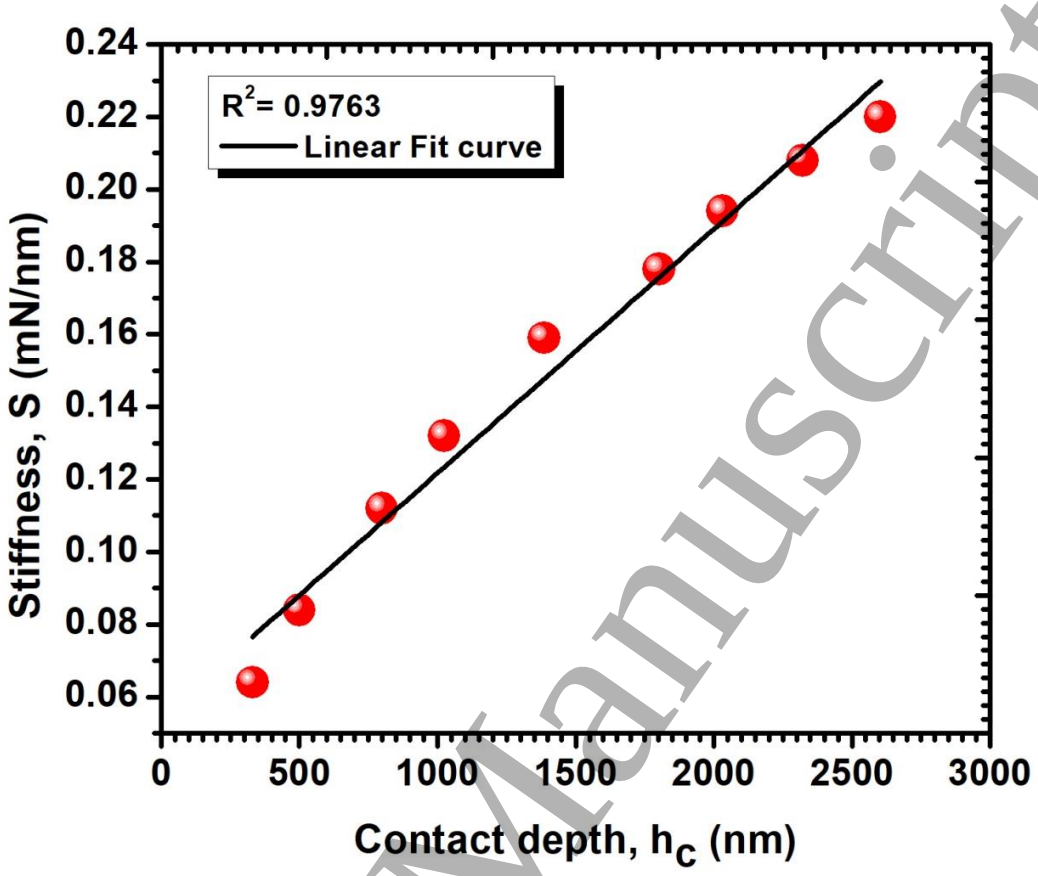


Fig. 9 Variation of stiffness with contact depth of LAHBr Single crystal

A linear relation can be observed between Stiffness and contact depth h_c in accordance with the relation

$$S = a + bh_c \qquad (8)$$

Where a and b correspond to indentation tip rounding and reduced Young's modulus respectively. A linear plot has been fitted to the curve and the slope of the fitted line gives the reduced Young's modulus. The reduced Young's modulus was found to be 6.733 $X10^{-5}$ mN nm^{-2} or 67.33 GPa for LAHBr single crystal.

Laser damage threshold (LDT):-

The laser damage threshold is one most important parameter that has to be considered when the system is integrating with optical components like mirrors. The Laser beam is highly monochromatic, directional and having intense energy which makes it capable to estimate the

Accepted Manuscript

damage sensitivity of the crystal. The laser damage threshold is a destructive technique that comprises two broad phenomena absorption-driven and dielectric-breakdown damage. The sample was exposed to the Nd:YAG laser of wavelength 1064 nm [21]. The damage occurs due to the laser was totally dependent upon the material properties which include absorption coefficient, specific heat, melting temperature as well as defects present in the crystal. Also, the geometrical properties like thickness, surface morphology and homogeneity can affect the efficiency of the crystal. In the present experiment, the damage threshold value has been examined for a plane (001) against the laser exposer. From the value of FWHM, it has been cleared that the crystal carries high crystalline perfection which in result gives a better value of damage threshold in comparison to its other family compound. As the defect concentration decreases, the energy absorption also decreases which leads to an increase in the value of laser-induced breakdown. Hence, this material proves to be an efficient candidate for optical applications. The damage threshold value for 1 pulse per second and 10 pulses per second is given below in Table 2.

Table 2 Laser damage threshold value for single and multiple shots

Crystal	Damage threshold			
	at 1 pulse per second		At 10 pulse per second	
LAHBr	31.6 J/cm^2	3.16 GW/cm^2	26.2 J/cm^2	2.62 GW/cm^2

These tabular values were further compared with L-arginine based compound for single pulse shot like L-arginine 4-nitrophenolate 4-nitrophenol dihydrate (LAPP) with a value of 2.54 GW/cm^2, L-arginine trifluoroacetate was 2 GW/cm^2 and L- arginine hydrochloride (LAHCL) was 0.011 GW/cm^2 [19,22,23]. By comparing the values, it was found that LAHBr has a higher value of LDT which was due to its high crystallinity as mentioned in HRXRD.

Shock damage threshold:-

A material may undergo physical and structural transformation due to sudden disturbances like an increase in pressure, temperature and mechanical stress. Such disturbances arise due to natural incidences such as earthquakes and volcanic eruption which lead to the generation of shock waves. Experimentally, the effect of these shock waves can be studied with the help of shock

Accepted Manuscript

tubes developed in a laboratory [24, 25]. For that purpose, a good quality LAHBr crystal of 2 mm thickness was placed at a distance of 1 cm from the open end of the driver section and then it was subjected, initially with low strength. After that, the strength was increased to Mach number 2.4. Before the experiment, the surface of the crystal was well polished and etched with water and the etched surface is shown in Fig. 10(a).

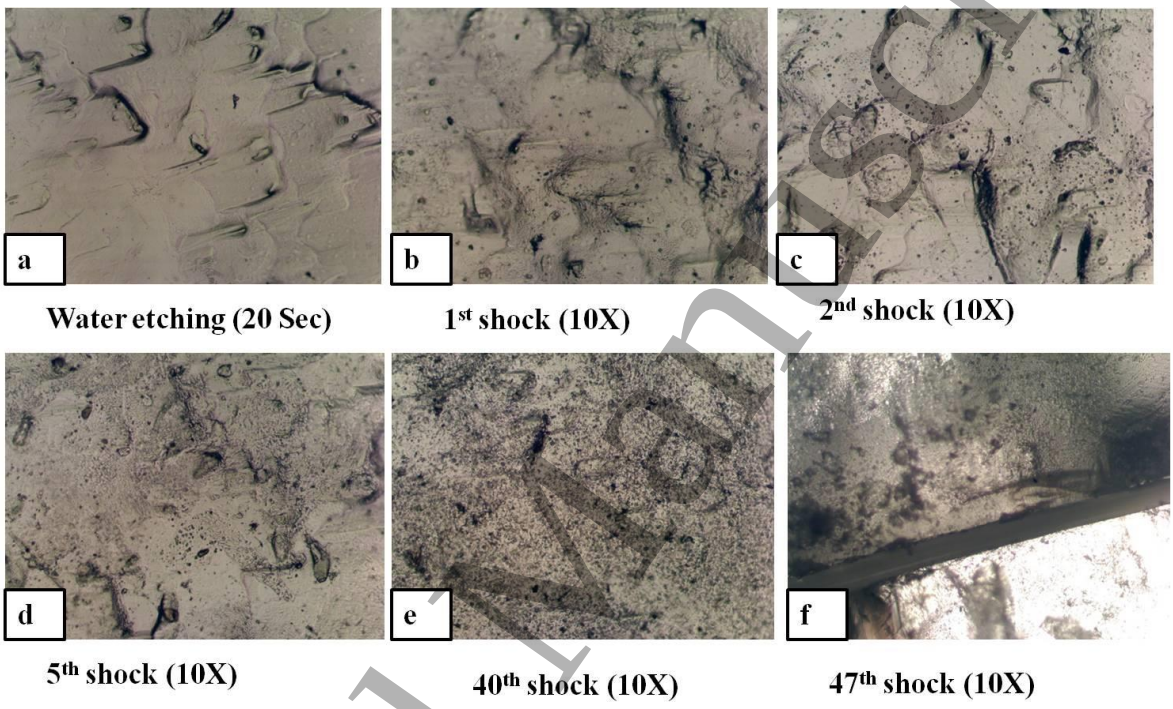


Fig. 10 Optical microscopy images of the crystal surface (a) before applied shocks; (b) after 1st shock; (c) after 2nd shock; (d) after 5th shock; (e) after 40th shock; (f) after 47th shock

Then the shock wave was exposed on the surface of the crystal and the surface was analyzed by an optical microscope as shown in Fig. 10(b). It is clear that the etch pattern is not affected and no crake is observed by the shock waves, but the molecules are peeled off from the surface which resulted as small circular pits. The pits are randomly scattered with irregular sizes. When the next shock waves are exposed, the pits densities are increased as the number of shocks increases in Fig. 10(c-e). Even for the 46th shock, the pit density reached the maximum but no crack was observed. But for the 47th shock, as the crystal was damaged with a crack (Fig. 10f) which is the shock damage threshold for the present crystal for Mach number 2.4. The results show that this crystal has very high shock resistivity and good mechanical stability. Hence this crystal can be used in the fast-moving and highly vibrating instruments.

Accepted Manuscript

Scanning Electron Micrsocope (SEM):-

The presence of defects on the surface of the specimen was examined by scanning electron microscope. The LAHBr single crystal was taken and then the surface was neatly polished by silk cloth and then subjected for the present study. The recorded SEM image on (001) direction is given in Fig. 11 (a) and (b). It is clearly visible that there is no remarakbale defects on the surface of the specimen In case if the surface is defected, the defected region is propogated inside the specimen and it will induce the microcracks. It will indirectly show grain boundaries. But in the present study, we have not noticed such type of defects. This is in tune with the recorded high resolution X-ray diffraction studies.

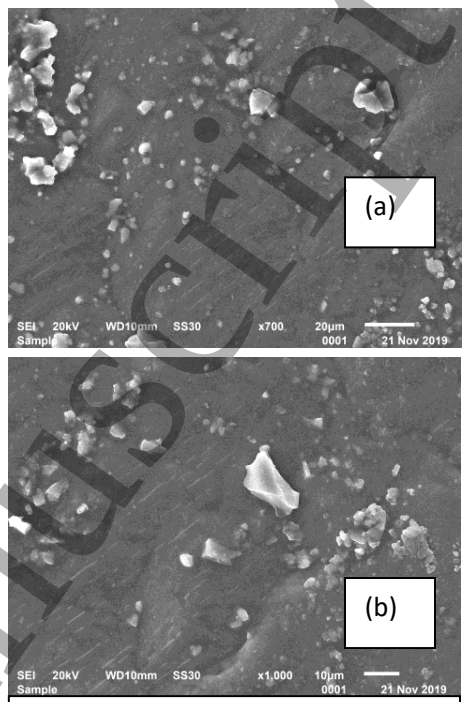


Fig. 11 (a) & (b)Surface image of LAHBr single crystal by SEM

Photoluminescence

Photoluminescence is an important technique for measuring the purity and crystalline quality of single crystals and for quantification of the amount of disorder present in them. Photoluminescence emission spectrum of the grown LAHBr crystal was recorded using an Edinburg luminescence spectrometer under ambient conditions. The recorded spectrum is shown in Fig. 12. Emission spectra of the single crystal LAHBr was recorded at an excitation wavelength of 235 nm. The LAHBR single crystal exhibit high intensity PL emission at 407 nm which indicates the titled crystal is a good candidate for absorption of ultraviolet light and emission of light in violet region. The less intense bands may be due to the lattice points in crystals, which are also being confirmed by the HRXRD analyses.

Accepted Manuscript

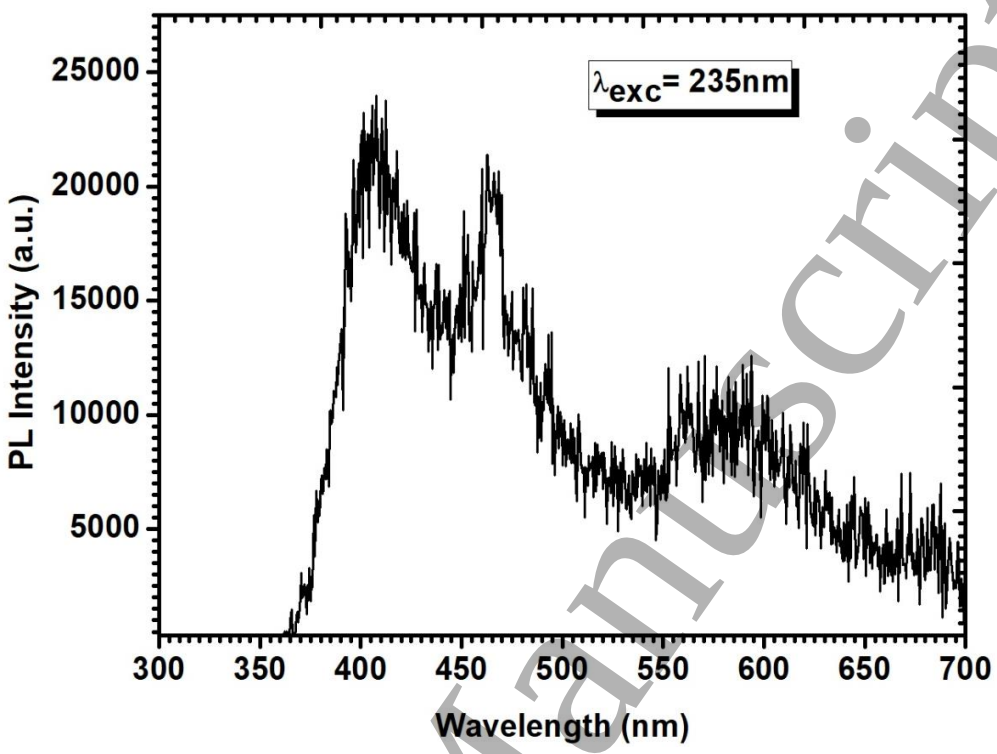


Fig. 12 Photoluminescence spectrum of LAHBr

Conclusion:-

A good quality single crystal has been grown by a modified way of unidirectional technique along (001) direction. The grown crystal carrying excellent crystalline perfection with the value of FWHM~9.31 arc second. Further, the crystal has been subjected to mechanical analysis. The titled compound was subjected to different loads and found high resistance against the deformation. Its resistivity towards the laser was evaluated using laser damage threshold technique which reveals that the crystal could be sustained for multiple shots. Its physical transformation was observed by applying shock waves to the crystal. The presence of defects on the surface was examined by SEM and it is in tune with the HRXRD technique.

Acknowledgments

The authors are very thankful to the Director, CSIR-NPL, for his continuous encouragement in carrying out the present study. One of the authors, Sonia is thankful to CSIR for giving financial assistance. Authors would like to acknowledge Prof. Subhasis Das, Dept. of Physics, University of Burdwan, Bardhaman, 713104, India for his guidance and support for LDT measurements.

Accepted Manuscript

Reference

1. Sonia, N. Vijayan, M. Vij, K. Thukral, N. Khan, D. Haranath, Rajnikant, M.S. Jayalakshmy, Investigation on the key aspects of L-arginine para nitrobenzoate monohydrate single crystal: A non-linear optical material, Chin. J. Chem. Eng. 27 (2019) 701-708.
2. M.D. Aggarwal, J. Stephens, A.K. Batra, R.B. Lal, Bulk crystal growth and characterization of semiorganic non-linear optical materials, J. Optoelectron Adv. M. 5 (2003) 555-562.
3. M. Senthil Pandian, U. Charoen In, P. Ramasamy, P. Manyum, M. Lenin, N. Balamurugan, Unidirectional growth of sulphamic acid single crystal and its quality analysis using etching, microhardness, HRXRD, UV–visible and Thermogravimetric-Differential thermal characterizations, J. Cryst. Growth. 312 (2010) 397–401.
4. M. Senthil Pandian, N. Balamurugan, G. Bhagavannarayana, P. Ramasamy, J. Cryst. Growth 310 (2008) 4143–4147.
5. M. Senthil Pandian, P. Ramasamy, Mater. Chem. Phys. 132 (2012) 1019–1028
6. X. Dong, J. Min-Hua, T. Zhong-Ke, A new phase matchable non-linear optical crystal-Larginine phosphate monohydrate (LAP), Acta Chim. Sin. 1 (1983) 230–233.
7. D. Xu, X.Q. Wang, W.T. Yu, S.X. Xu, G.H. Zhang, Crystal structure and characterization of a novel organic non-linear optical crystal: L-arginine trifluoroacetate, J. Cryst. Growth. 253 (2003) 481–487.
8. Z. Sun, G. Zhang, X. Wang, Z. Gao, X. Cheng, S. Zhang, D. Xu, Growth, Morphology, Thermal, Spectral, Linear, and Non-linear Optical Properties of L–arginine bis(trifluoroacetate) Crystal, Cryst. Growth Des. 9 (2009) 3251–3259.
9. V.V. Ghazaryan, B.A. Zakharov, A.M. Petrosyan, E.V. Boldyreva, L–argininium phosphite – a new candidate for NLO materials, Acta Crystallogr. Sect. C Struct. Chem. 71 (2015) 415–421.
10. L.N. Wang, X.Q. Wang, G.H. Zhang, X.T. Liu, Z.H. Sun, G.H. Sun, L. Wang, W.T. Yu, D. Xu, Single crystal growth, crystal structure and characterization of a novel crystal: L-arginine 4-nitrophenolate 4-nitrophenol dehydrate (LAPP), J. Cryst. Growth. 327 (2011) 133–139.
11. Jayesh R. Govani, William G. Durrer, Marian Manciu, Cristian Botez, and Felicia S. Manciu, J. Mater. Res., 2009, 24, 2316-2320.
12. S.K. Mazumdar, R. Srinivasan, Z. Kristallogr. 123 (1966) 186-205.

Accepted Manuscript

13. S. Mukerji, T. Kar, Structural, Thermal and Spectroscopic Investigation of non-linear Optical Crystal L-arginine Hydrobromide Monohydrate, Mater. Res. Bull. 33 (1998) 619–626.
14. B.W. Batterman, H. Cole, Dynamical Diffraction of X-rays by Perfect Crystals, Rev. Mod. Phys. 36 (1964) 681–717.
15. Sonia, N. Vijayan, M. Bhushan, K. Thukral, R. Raj, K.K. Maurya, D. Haranath, S.A. Martin Britto Dhas, Growth of a bulk-size single crystal of sulphamic acid by an in-house developed seed rotation solution growth technique and its characterization, J. Appl. Crystallogr. 50 (2017) 763–768.
16. M. Shkir, V. Ganesh, S. AlFaify, A. Black, E. Dieguez, K.K. Maurya, Large Size Crystal Growth, Photoluminescence, Crystal Excellence, and Hardness Properties of In-Doped Cadmium Zinc Telluride, Cryst. Growth Des. 18 (2018) 2046–2054.
17. W.C. Oliver, G.M. Pharr, An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments, J. Mater. Res. 7 (1992) 1564–1583.
18. I. Bdikin, B. Singh, J. Suresh Kumar, M.P.F. Graça, A.M. Balbashov, J. Grácio, A.L. Kholkin, Nanoindentation induced piezoelectricity in $SrTiO_3$ single crystals, Scr. Mater. 74 (2014) 76–79.
19. Sonia, N. Vijayan, M. Vij, P. Kumar, B. Singh, S. Das, R. Rajnikant, Assessment of the imperative features of an L-arginine 4-nitrophenolate 4-nitrophenol dihydrate single crystal for non-linear optical applications, Mater. Chem. Front. 1 (2017) 1107–1117.
20. Sonia, N. Vijayan, M. Vij, H. Yadav, R. Kumar, D. Sur, B. Singh, S.A. Martin Britto Dhas, S. Verma, Evaluation of structural, optical and mechanical behaviour of L-argininium bis(trifluoroacetate) single crystal: An efficient organic material for second harmonic generation applications, J. Phys. Chem. Solids. 129 (2019) 401–412.
21. N. Rani, N. Vijayan, B. Riscob, S. Karan Jat, A. Krishna, S. Das, G. Bhagavannarayana, B. Rathi, M.A. Wahab, Single crystal growth of ninhydrin by unidirectional Sankaranarayanan–Ramasamy (SR) method by using a glass ampoule for non-linear optical applications, Cryst. Eng. Comm. 15 (2013) 2127-2132.
22. X.J. Liu, Z.Y. Wang, X.Q. Wang, G.H. Zhang, S.X. Xu, A.D. Duan, S.J. Zhang, Z.H. Sun, D. Xu, Morphology and Physical Properties of L -arginine trifluoroacetate Crystals, Cryst. Growth Des. 8 (2008) 2270–2274.

Accepted Manuscript

23. A. Krishna, Sonia, N. Vijayan, S. Verma, B. Singh, I. Bidkin, M.S. Jayalakshmy, B. Sridhar, S. Das, Crystalline perfection, thermal, mechanical and optical investigations on solution grown L-arginine monohydrochloride single crystal, J. Mater. Sci. Mater. Electron. 28 (2017) 4306–4312.
24. G. Jagadeesh, Industrial applications of shock waves, J. Aero. Eng. 222 (2008) 575- 583.
25. G. Jagadeesh, K. Takayama, Novel applications of micro-shock waves in biological sciences, J. Indian Inst. Sci. 82 (2002) 49–57.

Accepted Manuscript

Check for updates

Article

Captivating Brand Hate Using Contemporary Metrics: A Structural Equation Modelling Approach

Vision
1–9
© 2020 MDI
Reprints and permissions:
in.sagepub.com/journals-permissions-india
DOI: 10.1177/0972262919892173
journals.sagepub.com/home/vis
SAGE

Richa Joshi[1]
Rajan Yadav[2]

Abstract

Brand polarity has gained considerable attention among researchers all over the world. Brands are trapped in a dilemma of love and hate experienced by their users. They are trying hard to place themselves in a favourable position which could fetch them more revenue and positive electronic word of mouth (e-WOM). Most of the literature in branding revolves around brand satisfaction, brand loyalty, brand equity and brand image, but there are relatively lesser studies which discuss brand hate (BH). However, it is a crucial emotion related to the brand. Therefore, this study has tried to fill this void in the literature by focusing on the antecedents of BH. The present research has tried to validate the effect of negative past experience (NPE) and subjective norm on BH along with the mediating effect of BH between the relationship of NPE and negative e-WOM. It has also tried to examine the moderating role of internet usage in the relationship of NPE and BH with negative e-WOM.

Key Words

Brand Hate, Mediation, Moderating Effect, Negative e-WOM, Negative Past Experience, Subjective Norm

Introduction

Brand polarity has gained popularity from the last few years, and scholars and practitioners are trying to capture the oscillation of human emotions from positive to negative towards various brands. Brand-associated behavioural aspect has got considerable attention, and such approach is found to be relatively new in capturing perspectives of consumers. However, positive emotions are discussed more in the existing researches. Brand love is found to be an intense feeling related to the brand, and various researchers such as Batra, Ahuvia, and Bagozzi (2012) and Carroll and Ahuvia (2006) have discussed this construct. Love is believed to be a deep emotion and so is hate; therefore it is vital to understand the underlying role of brand hate (BH) on the consumer perspective. People tend to remember and discuss negative aspects of the brand more than the positive ones. They discuss and share negative reviews more, whereas the positive experiences are less talked and discussed; it is termed as 'negativity bias' by Kanouse and Hanson (1972). Such incongruity paves a path for the researchers to explore the negative opinions and perspectives possessed by the consumer. With the advent of the internet, the expression of negative feelings is more frequent and easy these days. Social media platforms such as Twitter and Facebook are used for expression of opinion towards everything from politics to Bollywood and brands are not spared.

Though the construct carry such strong implications for marketers, still there is a dearth of studies which elaborate on BH in the literature of consumer behaviour. Hate is as important as love and the effect of BH on negative electronic word of mouth (e-WOM) is as significant for the marketer as the positive word of mouth. As per authors knowledge, there is no such study which explores the moderating role of internet usage in the relationships of BH and negative e-WOM. Along with this, the present study has

[1] Department of Management Studies, National Institute of Technology, Hamirpur, Himachal Pradesh, India.
[2] Delhi School of Management, Delhi Technological University, Delhi, India.

Corresponding author:
Dr. Richa Joshi, Assistant Professor, National Institute of Technology, Hamirpur, Himachal Pradesh 177005, India.
E-mail: richardias86@gmail.com

tried to exemplify a framework which accommodates the antecedents of BH and meditating role of BH in the relationship exhibited by negative past experience (NPE) and negative e-WOM.

Theoretical Grounding and Literature Review

In a study by Fehr and Russell (1984), it has been observed that after love, hate was found to be the most important emotion. Hate is said to be a complex feeling: It includes disgust, anger, fear, shame and disappointment (Zarantonello, Romani, Grappi, & Bagozzi, 2016). Though there exist fewer studies which discuss BH, but it is defined in various forms. It is expressed as an emotion (Hegner, Fetscherin, & van Delze, 2017; Weingarten, 2006; Zarantonello et al., 2016) in some studies, whereas other studies conceptualize it as an attitude (Ben, 2001). The dearth of literature for this construct has provided the authors enough motivation to explore its insights and its relationship with other constructs. The present study has proposed and empirically validated a framework (as shown in Figure 1) which demonstrates the effect of NPE and subjective norm on BH, along with the role of BH as a mediator in explaining the relationship of NPE and negative e-WOM. Unlike previous studies, negative word of mouth is adapted as negative e-WOM, owing to the spread of word of mouth via online sources. Additionally, internet usage is taken as moderator which influences the relationship of BH with negative e-WOM.

Negative Past Experience

It is the previous unpleasant experience of the brand which includes the functional failure and durability issues related with the product. Consumers have certain expectations before the usage of the product, and after using the product they compare the actual outcome received (Oliver, 1980). The difference in both can delight or dissatisfy the customer. When the performance is below the expectation, then it causes negative disconfirmation and can eventually lead to BH. Zarantonello et al. (2016) have also emphasized in their study that violation of expectations is the primary factor which causes BH among the customers. Hegner et al. (2017) have also shown the relationship of NPE and BH. Therefore, on the basis of the existing work, we can hypothesize that

H_1: NPE significantly influences BH.

Subjective Norm

Subjective norm is an interesting construct. It has its origin in theory of reasoned action (TRA). Ajzen and Fishbein (1980) proposed that consumer behaviour is explained by subjective norm and attitude. Subjective norm is termed as 'the pressure that society exerts on an individual' (Londono, Davies, & Elms, 2017). It is based on the behaviour which is considered appropriate by friends, family, neighbours and co-workers. It is a motivation to act according to the compliance of others. Sometimes an individual is motivated

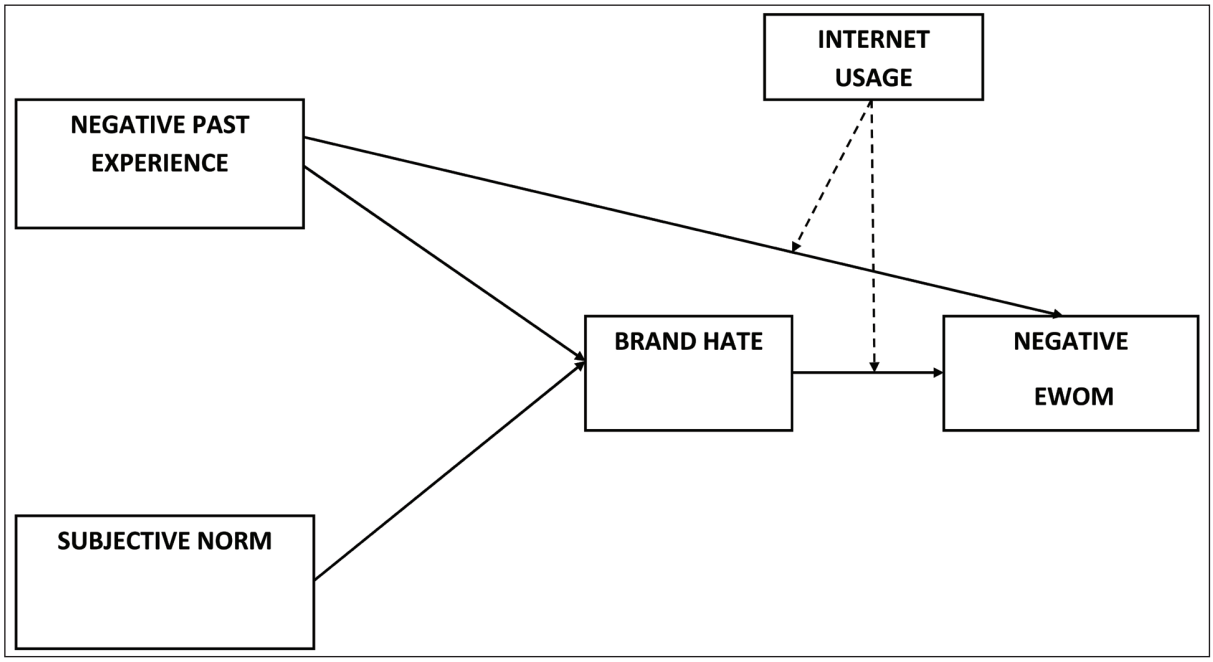


Figure 1. Research Framework

Source: The authors.

to act according to his attitude, but he may not perform that behaviour owing to its non-acceptance by the significant others (Hasbullah et al., 2016). Creating hatred/dislike towards a brand can also be constituted by the non-conformity of that brand by other people who are important for the consumer. Therefore, for the present framework, we hypothesize that

H_2: Subjective norm significantly influences BH.

Brand Hate

BH is conceptualized as a strong negative feeling towards the brand which is more than a mere dislike towards the brand. Consumers when feel frustrated with a brand show negative emotions where they tend to avoid those brands and sometimes they involve themselves in anti-branding activities. BH is classified into two types, that is, passive BH or attitudinal BH and active BH or behavioural BH. The former refers to a silent hate, whereas the latter refers to aggressively expressing the negative behaviour towards the brand (Kucuk, 2016, 2018; Zarantonello et al., 2016). While experiencing BH, consumers often end their relationships with the brand, and strong previous relationships increase the likelihood of more criticism of the brand. Such criticism is expressed many a times through online sources and is termed as e-WOM. E-WOM possess massive power these days. Since consumers today are empowered technologically, they can easily express the emotions felt by them by spreading their views on various blogs and social networking sites. Literature lacks the study on relationship of BH with negative e-WOM and NPE with negative e-WOM, though the relationships were explained with respect to negative word of mouth. Therefore, the present study has tried to fill this void and hence we can hypothesize that

H_3: NPE significantly influences negative e-WOM.
H_4: BH significantly mediates the relationship of NPE with negative e-WOM.

Negative Electronic Word of Mouth

Word of mouth is stated as 'all informal communications directed at other consumers about the ownership, usage, or characteristics of particular goods and services or their sellers' (Westbrook, 1987). The penetration of information technology has transformed its form to e-WOM where consumers are not only consulting their friends, family and acquaintances but also other unknown people who have used a particular product or service (Hennig-Thurau & Walsh, 2003). As it has been argued that 'bad things will produce larger, more consistent, more multifaceted, or more lasting effects than good things', when such informal information is negative it causes a lasting effect on the emotional front of the information seeker. However, it is vital to understand the underlying cause of such negative e-WOM which could be inherent in the hatred felt by the consumer.

Internet Usage as a Moderator

Internet usage has increased tremendously in the past five years, and with 460 million internet users India holds second position in its usage after China. Therefore, taking it as a moderating variable to study the relationship of NPE and BH with negative e-WOM was relevant for our study. Authors have measured internet usage on the basis of number of hours which helped in identifying the likelihood of spreading negative e-WOM if consumer had strong negative feeling or experience of the brand.

H_{5a}: The effect of NPE on negative e-WOM is moderated by internet usage.
H_{5b}: The effect of BH on negative e-WOM is moderated by internet usage.

Research Methodology

Data Collection

Data collection for the research was done using convenience sampling among the university students of various undergraduate and postgraduate courses. Respondents were advised to fill the questionnaire based on the hatred experienced by them concerning any cosmetic brand. Since consumers are highly sensitive to the choice of cosmetic brands, they express extreme emotions when their expectations are not fulfilled. Such expressions can be noticeable in the form of negative e-WOM. Questionnaire for the research consisted of three parts. The first part comprised of an open-ended question in which respondents had to state the most hated cosmetic brand by them and a brief reason for such hatred. The second part of the questionnaire included demographic information followed by the third portion which comprised of Likert scale-based questions. The research design was cross-sectional with the sample size of 374 responses out of 453 questionnaires send through offline and online mediums. Therefore, the response rate was 82.5 per cent.

Measurement Scales

Research consisted of four different constructs and the scale for framing the questionnaire was adopted from Hegner et al. (2017), Lada et al. (2009) as cited by Garg and Joshi (2018) and Salvatori (2007). However, scales were modified according to the requirements of the study. For estimating internet usage, the categories defined by Tiruwa, Yadav, and Suri (2018) (i.e., > 10 hours per week and < 10 hours per week) were used. The other four constructs were measured using Likert five point scale

(ranging from strongly disagree (1) to strongly agree (5)). Table 2 shows that the Cronbach alpha values for the constructs were more than the 0.7. Therefore, there was the presence of internal consistency in their respective items. Also, preliminary testing through exploratory factor analysis resulted in the clear demarcation of four distinct factors. The factor loadings for all items were more than 0.5 with a variance of 73.624 per cent. Therefore, each item loaded strongly on their respective construct/factor. Kaiser–Meyer–Olkin (KMO) test was used as a measure for estimation of sampling adequacy. The obtained value for the present study was 0.716 which was well above the prescribed threshold value by Tabachnick and Fidell (2001). Barlett's test of sphericity was also significant ($0.000 < 0.05$, $\chi^2 = 2{,}027.619$ and $df = 66$). This helps in acceptance of an innate alternative hypothesis in factor analysis, that is, variables are correlated in population (Joshi & Yadav, 2019).

Table 1. Demographic Details

Demographics	Frequency	Percentage
Gender		
Male	104	27.81
Female	270	72.19
Education		
Up to 12th	201	53.74
Graduation	173	46.26
Annual Family Income		
Less than 5 lakh	57	15.24
5.1 lakh –7.5 lakh	89	23.79
7.6 lakh –10 lakh	164	43.85
Above 10 lakh	64	17.12
Internet Usage		
> 10 hours per week	257	68.72
< 10 hours per week	117	31.28

Source: The authors.

Demographic Profile of Respondents

The respondents were the students of various postgraduate and undergraduate courses of Delhi University. The age of the respondents was ranging from 18 years to 25 years. Out of 374 total respondents, 104 were male and 270 were female. Large number of females were involved in the study since they are the prime users of cosmetic products. A total of 201 respondents were from graduation courses. Therefore, their highest qualification was up to 12th class, whereas 173 respondents were studying in postgraduate courses with their highest qualification as graduation. Different income group people were covered in the study. Around 57 respondents had family income less than 5 lakhs, 89 respondents belonged to the category of income range from 5.1 lakhs to 7.5 lakhs, 164 respondents had family income between 7.6 lakhs and 10 lakhs, and 64 respondents had family income greater than 10 lakhs. Out of the total respondents, 257 stated that they were using internet for more than 10 hours per week, whereas 117 respondents were using internet for less than 10 hours per week. Table 1 shows the demographic details of respondents.

Table 2. Reliability and Factor Loadings

Construct	Scale	Source	Cronbach α	Factor Loadings
Negative past experience	The performance brand X is poor.	Hegner et al. (2017)	0.787	0.863
	My hate for X brand is linked to the bad performance of this product.			0.844
	I am dissatisfied by brand X.			0.761
Subjective norm	Most people who are important to me think I should not purchase products of X brand when going for purchasing.	Lada et al. (2009) as cited by Garg and Joshi (2018)	0.845	0.881
	My family members do not prefer X brand.			0.866
	My friends would think that I should not choose X brand.			0.848
Brand hate	I am disgusted by brand X.	Hegner et al. (2017)	0.707	0.883
	I do not tolerate brand X and its company.			0.856
	Brand X is awful.			0.529
Negative electronic word of mouth	I spread negative word of mouth about brand X through online sources.	Hegner et al. (2017) modified for the study	0.899	0.888
	I try to influence a lot of people in not purchasing X brand via online sources.			0.901
	I try to spread my hate for X brand online.	Salvatori (2007) modified for the study		0.872
Internet usage	My internet usage per week (> 10 hrs per week/ < 10 hrs per week)	Tiruwa et al. (2018)	NA	

Source: The authors.

Table 3. Skewness and Kurtosis Statistics

Descriptive Statistics									
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
NPE	374	4.00	15.00	11.0294	2.49110	–0.435	0.126	–0.328	0.252
WOM	374	4.00	15.00	11.2379	2.71418	–0.756	0.126	0.099	0.252
BH	374	4.00	15.00	11.2861	2.19333	–0.404	0.126	–0.148	0.252
SN	374	3.00	15.00	10.6107	2.84834	–0.519	0.126	–0.457	0.252
Valid N (list-wise)	374								

Source: The authors.
Notes: Indices—NPE = negative past experience, WOM = negative word of mouth, BH = brand hate, SN = subjective norm.

Empirical Analysis

The analysis of the present study was done in several stages. The first stage comprised of testing the assumptions of structural equation modelling (SEM) to validate its applicability. The second stage consisted of confirmatory factor analysis (CFA) for verifying reliability and validity. The third stage of the analysis covered the empirical testing of all the hypotheses followed by the study of group differences concerning internet usage.

Testing the Assumptions of Structural Equation Modelling

In statistics, applicability of every technique is entirely dependent on the prerequisite of its assumptions. Hence, before application of the SEM, researchers had tested few assumptions. The sample size of the study was 374 which exceeded the required threshold of 200 (Hair, Black, Babin, & Anderson, 2013). Skewness and kurtosis values (as shown in Table 3) were $< \pm 3$ & $< \pm 10$ (Hair, Black, Babin, & Anderson, 2010), respectively. Therefore, assumption of normality was fulfilled. Linearity was estimated through curve estimation in regression which indicated that *f* values were higher and significant for the linear relationship in comparison to cubic, quadratic and inverse relationships (Joshi & Yadav, 2018).

Confirmatory Factor Analysis

Validity and reliability of various constructs were estimated in CFA. In this stage, constructs were allowed to freely correlate with each other (as shown in measurement model: Figure 2). They were estimated through the plug-ins proposed by Gaskin (2016). The results (Table 4) showed that the composite reliability values for all the constructs were more than 0.7. Also, the maximal reliability was > 0.7 which sufficed the requirements of the threshold suggested by Hair et al. (2013) and Kamata, Turhan, and Darandari (2003). For estimation of convergent validity, the values of average variance extracted (AVE) were examined. The values for all the constructs were more than 0.5 (Fornell & Larcker, 1981). Discriminant validity was also achieved since the square root of AVE is greater than inter-construct correlations (Fornell & Larcker, 1981).

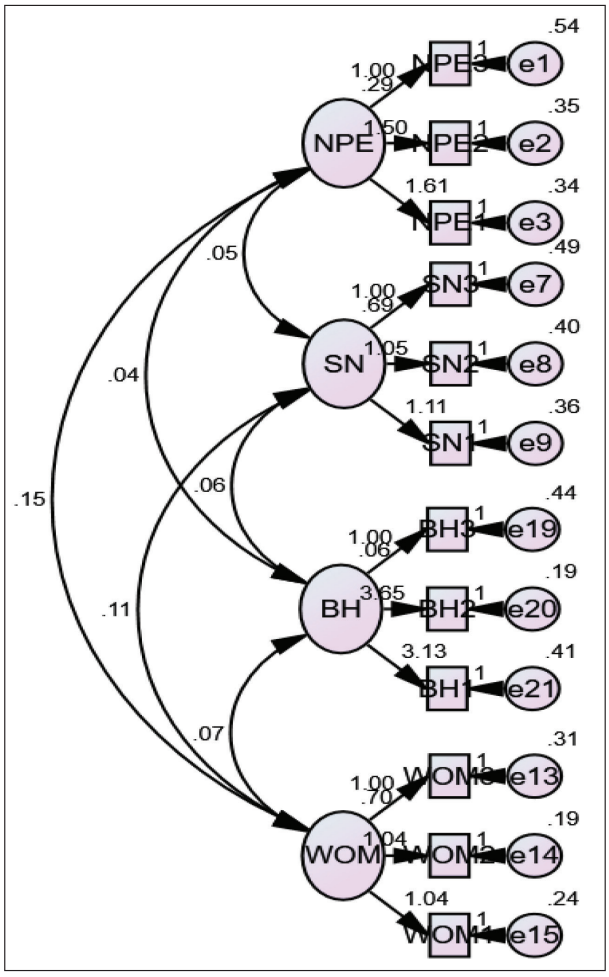


Figure 2. Measurement Model for CFA
Source: The authors.

Table 4. Reliability and Validity Estimates

	CR	AVE	MSV	MaxR(H)	NPE	SN	WOM	BH
NPE	0.794	0.567	0.105	0.825	**0.753**			
SN	0.846	0.646	0.077	0.849	0.117†	**0.804**		
WOM	0.900	0.750	0.116	0.903	0.322***	0.166**	**0.866**	
BH	0.737	0.512	0.116	0.856	0.324***	0.277***	0.340***	0.715

Source: The authors.

Notes: * denotes *p*-value is < 0.05 hence significant at 5% level of significance.

** denotes *p*-value is < 0.01 hence significant at 1% level of significance.

*** denotes *p*-value is < 0.001 hence significant at 0.1% level of significance.

Bold values indicate that square root of AVE (average variance extracted) is more than the inter-construct correlations hence discriminant validity was achieved.

Table 5. Model Fit Indices

X^2/df	GFI	CFI	NFI	TLI	RMSEA
3.049	0.941	0.951	0.929	0.932	0.074

Source: The authors.

Model Fit Indices

Model fit indices were evaluated with the help of prescribed thresholds in the previous studies. The results of our study, as shown in Table 5, indicates that $X^2/df = 3.049$ which is $\leq$ 3, goodness of fit index (GFI) is 0.941 which is $\geq$ 0.90. Comparative fit index (CFI), normed fit index (NFI), Tucker–Lewis index (TLI) have obtained the values of 0.951, 0.929 and 0.932, respectively, which are $\geq$ 0.90. Therefore, the model is a good fit (Hu & Bentler, 1999). However, the value of root mean square error of approximation (RMSEA) should be less, that is, smaller values indicate the better fit. The obtained value is 0.074 which is $\leq$0.08. Therefore, the model has an acceptable fit (Hair et al., 2010; Joshi & Yadav, 2017; Tiruwa et al., 2018).

Path Analysis and Hypotheses Testing

Path analysis of structural model (as shown in Figure 3) revealed that the *p*-values/sig values for all the proposed relationships were < 0.05 and 0.01 (as shown in Table 6). This led to the conclusion that all the hypotheses were accepted at both 5 per cent and 1 per cent level of significance. This indicates that NPE and subjective norm had a significant effect on BH. Therefore, H_1 and H_2 were accepted. Additionally, NPE had a significant effect on negative e-WOM. The effect was also mediated through BH. Therefore, it is a partial mediation effect and hence H_4 was also accepted.

Table 6. Results of Structural Analysis

S. No	Direct/Indirect Relationships	*p*-Value/Sig Value	Hypotheses Support	Relationship Tested
H_1	NPE ⟶ BH	0.000	Supported	Direct
H_2	SN ⟶ BH	0.001	Supported	Direct
H_3	NPE ⟶ WOM	0.002	Supported	Direct
H_4	NPE ⟶ BH⟶ WOM	0.000	Supported	Indirect

Source: The authors.

Moderating Effects

The study has utilized multi-group capabilities of SEM to study group differences (as shown in Figure 4) between the independent and dependent variables. Respondents were divided into two subgroups on the basis of their internet usage, that is, < 10 hours and > 10 hours. Moderating effect of internet usage was then evaluated among the relationships of NPE with negative e-WOM and BH with negative e-WOM. Results of multi-group shows that internet usage moderates both the relationships (Z = 3.433*** and Z = –3.31*** which is significant at 1 per cent level of significance as shown in Table 7). Therefore, H_{5a} and H_{5b} were accepted.

Discussion and Conclusion

The present research article has contributed to the literature of emotional connections with the brands through its focus on an intense feeling with the brand, that is, BH. It is one of the few studies which are based on the quantitative analysis of BH. Results of the study show that NPE and subjective norm have a significant effect on BH. Additionally, the study has incorporated the role of technology while spreading negative word of mouth through the addition of negative e-WOM in the existing framework. Therefore, public complaining aspect of negative WOM is taken into account in the present study. Due to the proliferation of social media, it is easy to spread negative word of mouth electronically. People find it easier to write/discuss their opinions in the form of reviews, ratings and tweets. Hence, taking e-WOM in this research has given a fresh perspective to the existing literature. Hegner et al. (2017) mentioned that future studies could explore the role of moderator in their framework. Hence, the addition of internet usage as a moderator has strengthened the findings through the evaluation of group differences between two different set of internet users (less/more than 10 hours per week). The findings show that the effect of BH and NPE on negative word of mouth significantly differ for these two categories of internet users (i.e., less than 10 hours and more than 10 hours). A significant

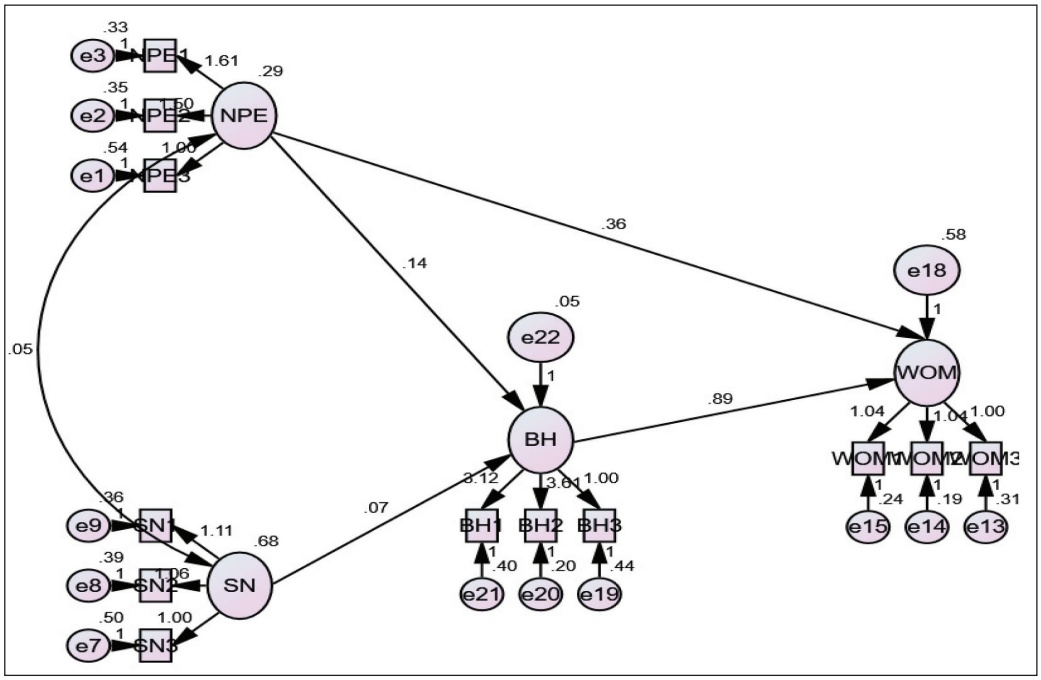


Figure 3. Structural Equation Model Stating Causal Relationships

Source: The authors.

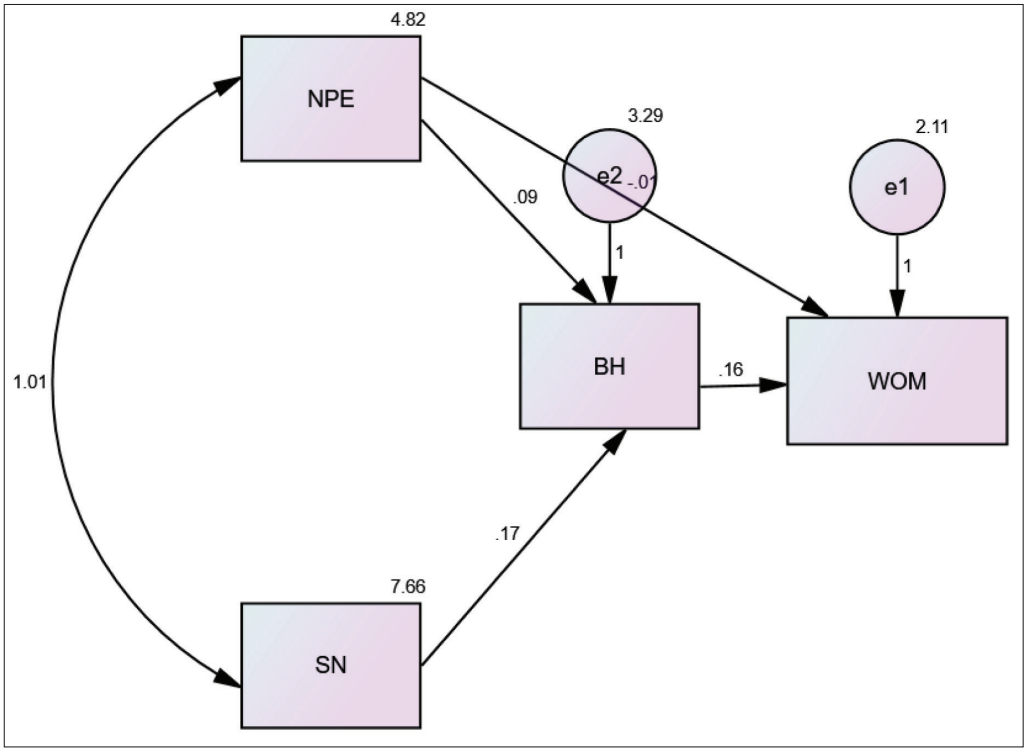


Figure 4. Moderating Effects of Internet Usage

Source: The authors.

Table 7. Moderating Effects of Internet Usage

S. No				More Than 10 Hours		Less Than 10 Hours			Hypotheses Support
				Estimate	*P*	Estimate	*p*	*z*-score	
H_{5a}	WOM	←	NPE	–0.008	0.852	0.284	0.000	3.433***	Supported
H_{5b}	WOM	←	BH	0.160	0.000	–0.185	0.045	–3.31***	Supported

Source: The authors.
Notes: *** denotes *p*-value < 0.01 hence significant at 1% level of significance.

direct effect of NPE on negative e-WOM has also validated the fact that it is not necessary for the users to hate the brand first before spreading any negative WOM/e-WOM. Rather sometimes NPE is immediately followed by the spread of negative e-WOM without any intense hatred feel towards the brand. Since cosmetic products belong to a higher involvement category, when any cosmetic brand fails to provide appropriate result the consumer immediately writes negative e-WOM on the websites and mobile applications such as Nykaa, Flipkart and Amazon.

Managerial Implications

Findings of the study shows that consumers are highly sensitive for their choice of cosmetic brands. Therefore, they are relatively quicker forming opinions and spreading WOM. It was also observed that subjective norm is one of the significant precursors of BH since consumer may develop feelings of hatred towards a brand due to the influence of significant others. One of the good examples is of a leading cosmetic brand which promotes fairness, hence few consumers in India hate that brand. Therefore, firms could emphasize on promoting products in such a way that it does not hurt sentiments of masses as consumers these days frame their purchase intention on the basis of social conformity. Usage of internet as a moderator has a strong implication for marketers. It states that consumers who are using internet for more than 10 hours a week are likely to spread negative WOM online through various ways. Therefore, brands should strive to have their presence online. They should also appoint few individuals who would serve as brand advocates to regularly keep a check on the posts, comments and reviews on application and websites of brands along with Facebook brand pages. They should immediately reply to the problems and dissatisfaction faced by the consumer concerning the brand.

Limitations and Future Research

Present research suffers from a few limitations such as the use of cross-sectional research design is unable to capture the changing paradigm of human emotions. Therefore, future researches can utilize a longitudinal research design to understand the impact of product failures on brand emotions. The usage of relatively small sample size is another limitation which could inhibit the generalization of results. Future researches can also conduct studies in other various countries of Asia and western parts of the world to have a holistic view of consumer emotions. Additionally, in this study, consumers were not provided any brand as a stimulus. Hence, in few instances consumers were unable to recall any cosmetic brand they hate. Studies in near future can take a set of 50–100 brands as a stimulus to generate a more quick response from the consumer.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Richa Joshi https://orcid.org/0000-0003-2262-0852

References

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice Hall.

Batra, R., Ahuvia, A., & Bagozzi, R. P. (2012). Brand love. *Journal of Marketing*, *76*(2), 1–16.

Ben-Ze'ev, A., & Ben-Ze'ev, A. (2001). *The subtlety of emotions*. MIT press.

Carroll, B. A., & Ahuvia, A. C. (2006). Some antecedents and outcomes of brand love. *Marketing Letters*, *17*(2), 79–89.

Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General*, *113*(3), 464.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50.

Garg, P., & Joshi, R. (2018). Purchase intention of 'halal' brands in India: The mediating effect of attitude. *Journal of Islamic Marketing*, *9*(3), 683–694.

Gaskin, J. & Lim, J. (2016). "Model Fit Measures", AMOS Plugin. Gaskination's StatWiki.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010), *Multivariate data analysis: A global perspective* (7th ed.). Englewood Cliffs, NJ: Prentice Hall.

———. (2013). *Multivariate data analysis: Pearson new international edition*. Essex, U.K: Pearson Education.

Hasbullah, N. A., Osman, A., Abdullah, S., Salahuddin, S. N., Ramlee, N. F., & Soha, H. M. (2016). The relationship of attitude, subjective norm and website usability on consumer intention to purchase online: An evidence of Malaysian youth. *Procedia Economics and Finance*, *35*, 493–502.

Hegner, S. M., Fetscherin, M., & van Delzen, M. (2017). Determinants and outcomes of brand hate. *Journal of Product & Brand Management*, *26*(1), 13–25.

Hennig-Thurau, T., & Walsh, G., (2003). Electronic word-of-mouth: Motives for and consequences of reading customer articulations on the Internet. *International Journal of Electronic Commerce, 8*(2), 51–74.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Joshi, R., & Yadav, R. (2017). Evaluating the feedback effects of brand extension on parent brand equity: A study on Indian FMCG industry. *Vision*, *21*(3), 305–313.

———. (2018). Exploring the mediating effect of parent brand reputation on brand equity. *Paradigm*, *22*(2), 125–142.

———. (2019). The study of brand extension among generation Y in the Indian market. *International Journal of Indian Culture and Business Management*, *19*(3), 339–353.

Kamata, A., Turhan, A., & Darandari, E. (2003, April). *Estimating reliability for multidimensional composite scale scores*. Paper presented at the annual meeting of American Educational Research Association, Chicago, IL.

Kanouse, D. E., & Hanson, L. R., Jr. (1972). Negativity in evaluations. In E. R. Jones et al. (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, NJ: General Learning Press.

Kucuk, S. U. (2016). What is brand hate? In *Brand hate* (pp. 17–36). Cham: Palgrave Macmillan.

———. (2018). Macro-level antecedents of consumer brand hate. *Journal of Consumer Marketing*, *35*(5), 555–564.

Londono, J. C., Davies, K., & Elms, J. (2017). Extending the theory of planned behavior to examine the role of anticipated negative emotions on channel intention: The case of an embarrassing product. *Journal of Retailing and Consumer Services*, *36*, 8–20.

Oliver, R. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research, 17*, 460–469.

Salvatori, E. (2007). *Brand hate: The "dark side" of consumer attitudes towards a brand* (Unpublished master's thesis). Milan: Bocconi University.

Tabachnick, B. G., & Fidell, L. S. (2001). *SAS for Windows workbook for Tabachnick and Fidell using multivariate statistics*. New York, NY: Allyn and Bacon.

Tiruwa, A., Yadav, R., & Suri, P. K. (2018). Modelling Facebook usage for collaborative learning in higher education. *Journal of Applied Research in Higher Education*, *10*(3), 357–379.

Weingarten, K. (2006). On hating to hate. *Family Process*, *45*(3), 277–288.

Westbrook, R. A. (1987). Product/consumption-based affective responses and postpurchase processes. *Journal of Marketing Research*, *24*(3), 258–270.

Zarantonello, L., Romani, S., Grappi, S., & Bagozzi, R. P. (2016). Brand hate. *Journal of Product & Brand Management*, *25*(1), 11–25.

About the Authors

Dr Richa Joshi (richardias86@gmail.com) is an Assistant Professor in National Institute of Technology (NIT) Hamirpur(H.P). She has more than 10 years of teaching experience. The areas of her interest are brand equity, brand extension and consumer brand relationships. She has completed her doctoral degree from Delhi Technological University. She has obtained her graduation from Delhi University and post-graduation from GGSIP University with the qualification of UGC-NET in Management. She has presented papers in various national and international conferences at IIM Indore, IIM Kashipur, MDI Gurgaon and IIT Delhi. She has contributed 'ABDC' and 'Scopus' indexed research papers in several journals of Sage, Inderscience and Emerald publications. She has also been an active reviewer of journals of Sage and Emerald.

Professor Rajan Yadav (raj_yadav1974@yahoo.co.in) is a Professor and the Head of Delhi School of Management, Delhi Technological University. His area of interest is marketing management. He graduated from the Delhi University and obtained his Master's in Business Management from the Maharshi Dayanand University. He was awarded PhD in Organized Retail and also qualified National Eligibility Test (NET) conducted by UGC. His research interests include behavioural dimensions of technology diffusion, online social networks, retail management and student community-related social marketing issues. He has 17 years of postgraduate teaching experience. He has contributed many research papers in journals of repute published by SAGE Publications, Emerald and Inderscience.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/338344602

Comparative Study of Improvements in Multiple Effect Vapor Absorption Systems using Loop Heat Pipes

Article · January 2020

DOI: 10.22214/ijraset.2019.12162

CITATIONS
0

READS
11

2 authors, including:

Ankit Dwivedi
Delhi Technological University
10 PUBLICATIONS **6** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Doctoral Research View project

All content following this page was uploaded by Ankit Dwivedi on 02 January 2020.

The user has requested enhancement of the downloaded file.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

Comparative Study of Improvements in Multiple Effect Vapor Absorption Systems using Loop Heat Pipes

Ankit Dwivedi[1], R S Mishra[2]
[1, 2]*Dept. of Mechanical Engineering, Delhi Technological University, Delhi*

Abstract: ***The research work is based on the research done on the modification of vapor absorption refrigeration systems (VARS) by incorporating Loop Heat Pipes (LHP). A comparison of the performance enhancements of different effect systems namely single, double, half and triple effect systems is performed in this research work. The modification in the systems is brought in by replacing the condenser of the VARS systems by LHPs. It's shown through the simulations that intra cycle heat exchange increases with the change in design, materials of the LHP and operating temperatures. The COP $_{I}$ and COP $_{II}$ are maximum for triple effect system whereas the percentage increase in the performance is highest in double effect system 78% & 28%. Higher the generator temperature, higher is the heat utilization factor for the LHPs. The system becomes concise and flexible at the same time.***

Keywords: Loop Heat Pipes, Vapor Absorption Refrigeration System, Half Effect, Double Effect, Triple Effect, Double Effect

I. INTRODUCTION

The performance parameters of different Vapor absorption systems have been studied. The systems have been modified by removal of condenser from the original system. The advantages in the new systems have been discussed earlier, this research work happens to be comparing all the aspects.

II. LITERATURE REVIEW

Fabian Korn et al. [2012] performed several vital experiments on heat pipes to establish it to be one of the most effective procedures to transport thermal energy from one point to another, mostly used for cooling[1]. R.Z. Wang et al.[2008]added heat pipes in adsorption water chiller or ice maker initials. His work showed that the adsorption refrigerators are very efficient [2]..R. Rajashree et al.[1990]went through a numerical analysis of an unsteady, viscous, laminar, incompressible, two dimensional heat and mass transfer, in the vapour gas region of gas loaded circular heat pipe [3]. M.M. Talbi et al. (2000) carried out an exergy analysis on a single-effect absorption refrigeration cycle with lithium-bromide±water as the working Fluid pair [4]. E. Kurem et al.(2001) analyzed the Absorption Heat Pump (AHP) and Absorption Heat Transformers (AHT) using ammonia-water and water-lithium bromide solutions. A fundamental AHP and AHT systems was described and explained the operating sequence [5]. S.A. Adewusi et al (2004). studied the performance of single-stage and two- stage ammonia–water absorption refrigeration systems (ARSs).They calculated entropy generation of each component and the total entropy generation of all the system components as well as COP of the ARSs [6] . S. Arivazhagan et al. (2006) investigated experimentally on the performance of a two-stage half effect vapour absorption cooling system .The prototype is designed for 1 kW cooling capacity using HFC based working fluids (R134a as refrigerant and DMAC as absorbent) [7] . Rabah Gomri et al. (2008) performed exergy analysis of double effect lithium bromide/water absorption refrigeration system. The system consisted of a second effect generator between the generator and condenser of the single effect absorption refrigeration system, including two solution heat exchangers between the absorber and the two generators [8]. S.C. Kaushika et al. (2009) presented the energy and exergy analysis of single effect and series flow double effect water–lithium bromide absorption systems. They developed a computational model for the parametric investigation of the systems [9]. Gulshan Sachdeva et al.(2014) performed an exergy analysis of VAR system using LiBr- H2O as working fluid with the modified Gouy-Stodola approach [10]. Ankit Dwivedi et al. (2018) through computer simulations of the replacement of condensers from single, double, half &triple effect VARS, showed the improvements in the performance in COP $_{I}$ and COP $_{II}$ [11].
Based on the research work presented in the section, the comparison of the improvements in the performance parameters in all the different type of VARS based on the effect is due. This research article focuses on the comparative analysis. It will be described how the COPI and COPII vary for across the VAR systems with heat interaction and temperature in the LHP.

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

III. SYSTEMS DESCRIPTION

The following Fig.1 Shows the modification made by Ankit Dwivedi et. al. [42] in the simple single effect VARS operating between 373K & 273K . The system is able to reduce the waste of heat very effectively. Fig. Shows the changes made by Ankit Dwivedi et. al. [44] in Double effect VARS operating between 403K and 277K . The system operates without the presence of condenser. Also the heat exchange has improved with the use of LHP between low temperature generator and absorber. The LHP works on a very constant temperature, that reduces the exergy loss due to heat transfer through a large temperature difference.

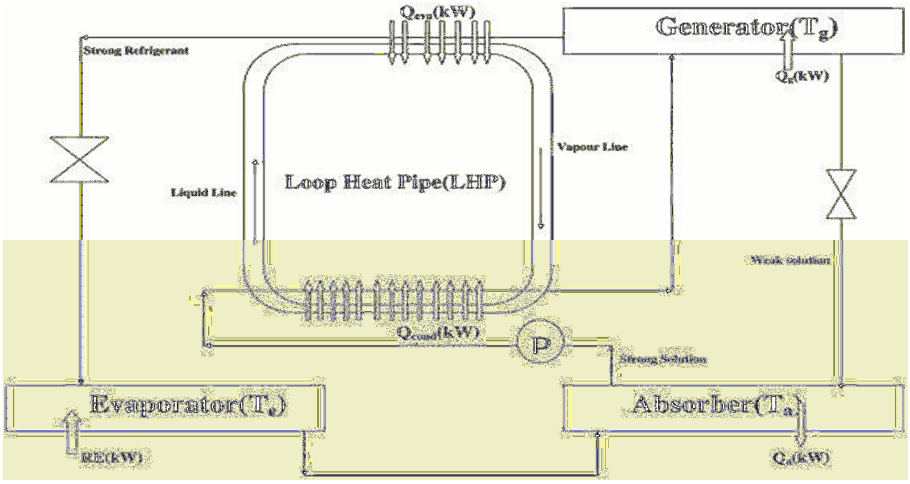


Fig.1: Modified VARS with a LHP [42].

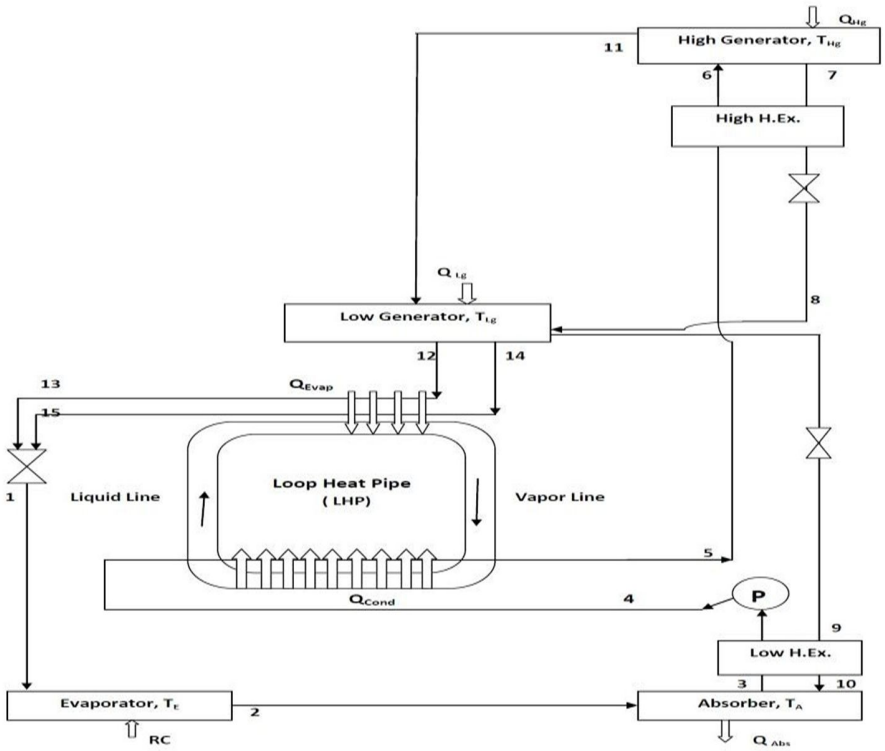


Fig.2: Modified Double Effect VARS with a LHP [44]

Fig. 2 is the schematic of the modifiation of Half effect VARS by installing the LHP arranging heat exchange between High Generator and High Absorber, for the similar reason of reducing heat input to the generator and making possible of effective utilization of heat from low temperature sources. The system operates between 363K & 278K.

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

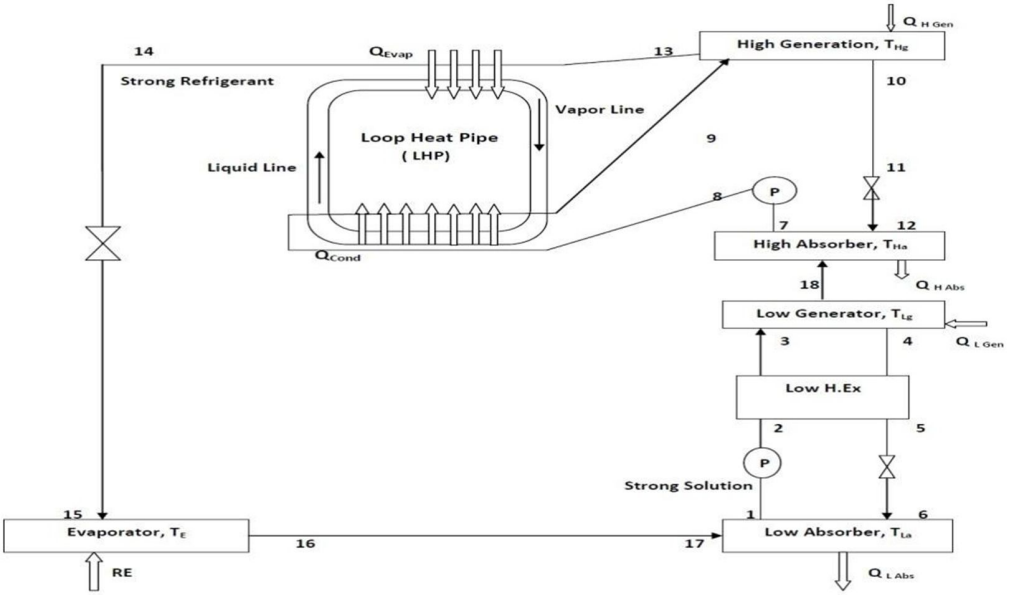


Fig.3: Modified Half Effect VARS with a LHP [43]

Triple effect system is modified by using 2 LHPs as can be seen in the Fig. 3 Mixture from the absorber is pumped directly through the LHPs. With the help of these LHPs, heat exchange at two levels is arranged. This system operates between 433K and 276 K. These designed systems in Fig. 4 are obtained by removing condensers by LHPs. The design is flexible and intra-cycle heat exchange is achieved, heat loss to the environment is reduced, reducing the entropy generation.

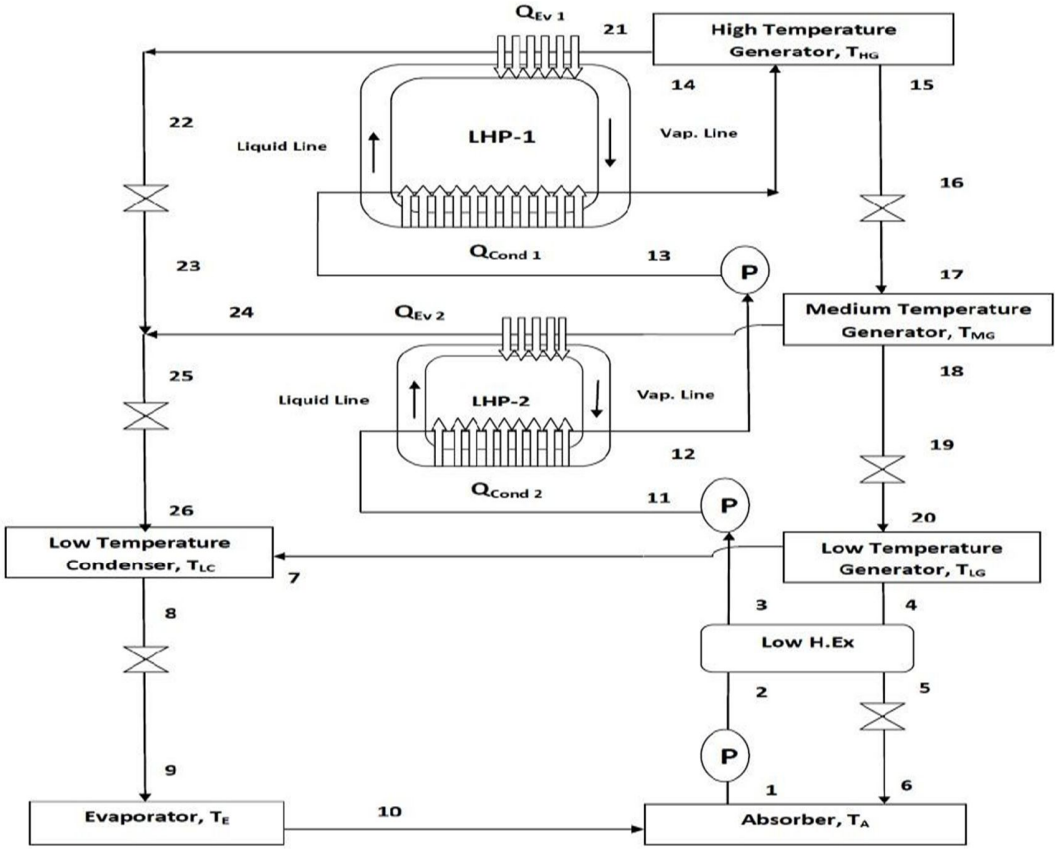


Fig.4: Modified Tripple Effect VARS with a LHP [45]

©IJRASET: All Rights are Reserved

The following tables contain the parameters used in the systems explained above in order to obtain the required simulations.

Table1: Terms Used in Simulation of Single Effect VARS [46].

Terms	**Abbreviations**
Refrigeration Effect in kW	RE (kW)
Heat rejected in absorber in kW	Q_a (kW)
Heat supplied in generator in kW	Q_g (kW)
Heat rejected in condenser of LHP in kW	Q_{cond} (kW)
Heat absorbed in evaporator of LHP in kW	Q_{eva} (kW)
Absorber Temperature in K	T_a,T_A (K)
Generator Temperature in K	T_g, T_G (K)
Condenser Temperature in K	T_c (K)
Evaporator Temperature in K	T_E,T_e (K)
Heat Rejected in Condenser in kW	Q_C (kW)
First Law Coefficient of Performance	COP_I
Second Law Coefficient of Performance	COP_{II}
Heat Leaked from the LHP in kW	Q_{Leak} (kW)
Percentage Improvement in First Law Coefficient of Performance	$\%COP_{I\ imp}$
Percentage Improvement in Second Law Coefficient of Performance	%COP II imp
Improvement in First Law Coefficient of Performance	COP I imp
Improvement in Second Law Coefficient of Performance	COP II imp

Table2: Terms Used in Simulation of Double Effect VARS[44]

Terms	Abbreviations
Refrigeration Effect in kW	RE (kW)
Heat rejected in absorber in kW	Q_a (kW)
Heat supplied in generator in kW	Q_g (kW)
Heat rejected in condenser of LHP in kW	Q_{cond} (kW)
Heat absorbed in evaporator of LHP in kW	Q_{eva} (kW)
Absorber Temperature in °C	T_{La},T_{Ha} (°C)
Generator Temperature in °C	T_{Hg}, T_{LG} (°C)
LHP Condenser Temperature in °C	T_c (°C)
Evaporator Temperature in °C	T_E,T_e (°C)
Heat Rejected in Condenser in kW	Q_C (kW)
First Law Coefficient of Performance	COP_I
Second Law Coefficient of Performance	COP_{II}
Heat Leaked from the LHP in kW	Q_{Leak} (kW)
Percentage Improvement in First LawCoefficient of Performance	$\%COP_{I\ imp}$
Percentage Improvement in Second LawCoefficient of Performance	%COP II imp
Improvement in First Law Coefficient ofPerformance	COP I imp
Improvement in Second Law Coefficient of Performance	COP II imp
Low Pressure Generator	LPG
High Pressure Generator	HPG

©IJRASET: All Rights are Reserved

Table3: Terms Used in Simulation of Half Effect VARS [43]

Terms	**Abbreviations**
Refrigeration Effect in kW	RE (kW)
Heat rejected in absorber in kW	Q_a (kW)
Heat supplied in generator in kW	Q_g (kW)
Heat rejected in condenser of LHP in kW	Q_{cond} (kW)
Heat absorbed in evaporator of LHP in kW	Q_{eva} (kW)
Absorber Temperature in °C	T_{La},T_{Ha} (°C)
Generator Temperature in °C	$T_{Hg,}$ T_{LG} (°C)
LHP Condenser Temperature in °C	T_c (°C)
Evaporator Temperature in °C	T_E,T_e (°C)
Heat Rejected in Condenser in kW	Q_C (kW)
First Law Coefficient of Performance	COP_I
Second Law Coefficient of Performance	COP_{II}
Heat Leaked from the LHP in kW	Q_{Leak} (kW)
Percentage Improvement in First Law Coefficient of Performance	$\%COP_{I\ imp}$
Percentage Improvement in Second Law Coefficient of Performance	%COP II imp
Improvement in First Law Coefficient of Performance	COP I imp
Improvement in Second Law Coefficient of Performance	COP II imp

Table4: Terms Used in Simulation of Triple Effect VARS [45]

Terms	Abbreviations
Refrigeration Effect in kW	RE (kW)
Heat rejected in absorber in kW	Q_a (kW)
Heat supplied in generator in kW	Q_g (kW)
Heat rejected in condenser of LHP in kW	Q_{cond} (kW)
Heat absorbed in evaporator of LHP in kW	Q_{eva} (kW)
Absorber Temperature in °C	T_{La},T_{Ha} (°C)
Generator Temperature in °C	$T_{Hg,}$ T_{LG},T_{MG} (°C)
LHP Condenser Temperature in °C	T_c (°C)
Evaporator Temperature in °C	T_E,T_e (°C)
Heat Rejected in Condenser in kW	Q_C (kW)
First Law Coefficient of Performance	COP_I
Second Law Coefficient of Performance	COP_{II}
Heat Leaked from the LHP in kW	Q_{Leak} (kW)
Percentage Improvement in First Law Coefficient of Performance	$\%COP_{I\ imp}$
Percentage Improvement in Second Law Coefficient of Performance	%COP II imp
Improvement in First Law Coefficient of Performance	COP I imp
Improvement in Second Law Coefficient of Performance	COP II imp
Low Temperature Generator, Condenser	LG,LC
Medium Temperature Generator, Condenser	MG, MC
High Temperature Generator, Condenser	HG, HC

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

IV. RESULTS AND DICSUSSIONS

Based on the simulations this section revolves abounds the results obtained. The COP_I, COP_{II} are analysed with variations in Q_{Cond} & T_C . The performance enhancements in the above detailed systems are compared.

A. Comparison of Improvements in COP_I & COP_{II} with Varying Heat Input in Condenser of the Loop Heat Pipe (LHP)

This section is based on the results obtained by varying the parameter Heat Exchange in the condenser of the LHP. It's the energy that moves from the condensing strong refrigerant in the evaporator of the LHP. The idea is to transfer this energy to the mixture out of the absorber, to reduce the heat loss, improving both the COPs. Fig shows the COPs of single effect.

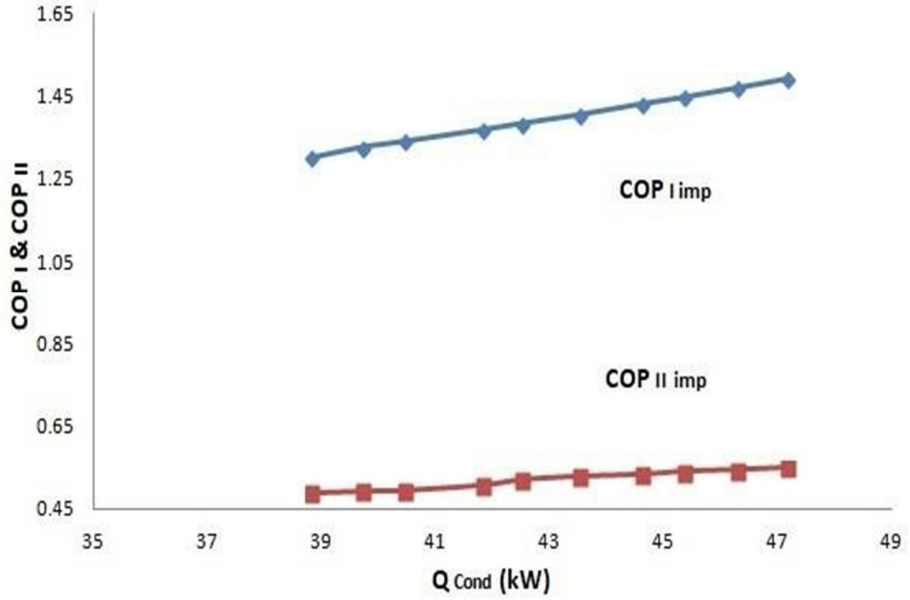


Fig.5: Single Effect

Fig is the plot of COPs with Q_{Cond} for the double effect system, whereas Fig is showing results for half effect system. The results of triple effect follows in fig .

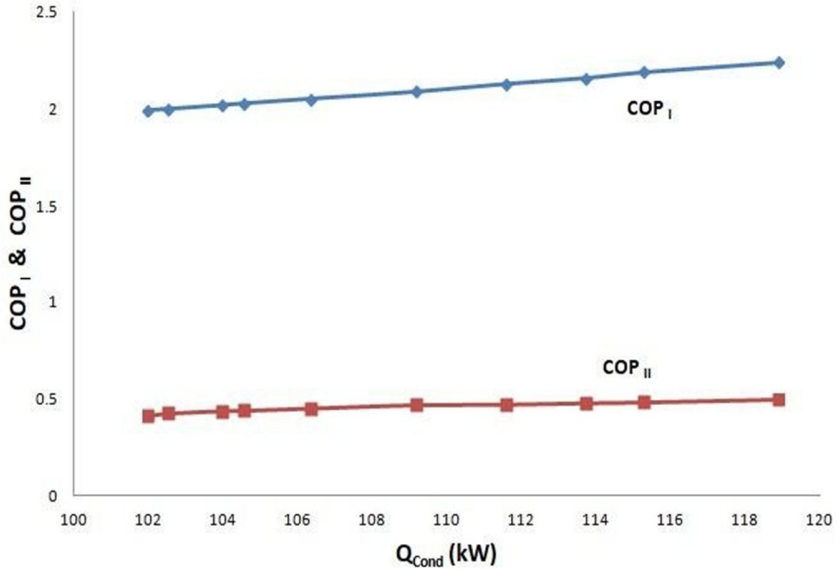


Fig.6: Double Effect

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

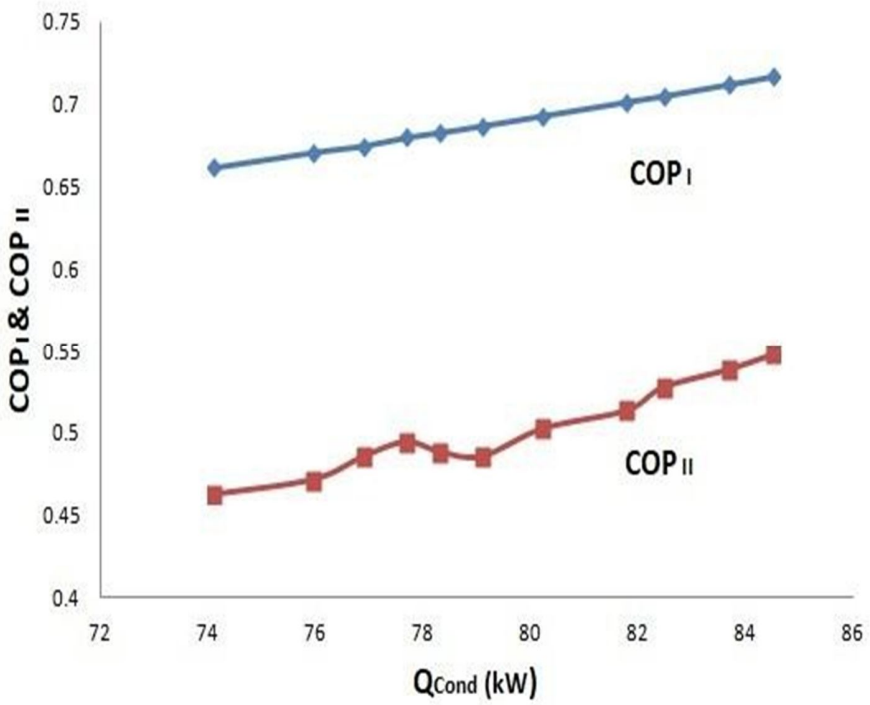


Fig.7: Half Effect

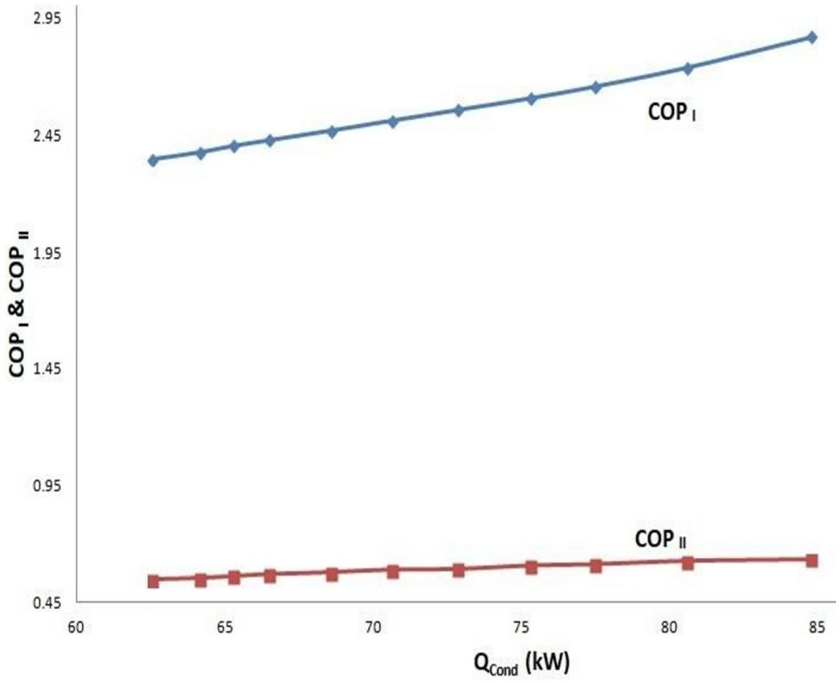


Fig.8: Triple Effect

It can be observed from the above figures that average COP_I & COP_{II} is for the improved single effect system are 1.4 & 0.51respectively. For the double effect system the figures of average COP_I & COP_{II} are 2.09 & 0.4571 respectively. The modified half effect system has average COP_I & COP_{II} as 0.69 and 0.502 respectively. The enhanced average COP_I & COP_{II} for Triple effect system is found to be 2.545 and0.588.

©IJRASET: All Rights are Reserved

B. Comparison of Percentage Improvements in COP_I & COP_{II} with Varying Heat Input in Condenser of the Loop Heat Pipe (LHP)

The above section explains the values of the COPs for the modified systems, while in this section it will be clear about the percentage enhancement of the COPs compared to that of the conventional systems with varying Q_{Cond}. Fig shows the average improvements in COP_I & COP_{II} for modified single effect system to be 68% and 23% respectively. Moreover Fig if showing the results for modified double effect system where average enhancement in COP_I & COP_{II} is 75.85% & 28.5 % respectively. Average increase in COP_I & COP_{II} is 64% & 27 % for the modified half effect system. The new triple effect system gets the improvement in COP_I & COP_{II} of about 61% and 25% as can be accessed from the Fig 9 to fig 12.

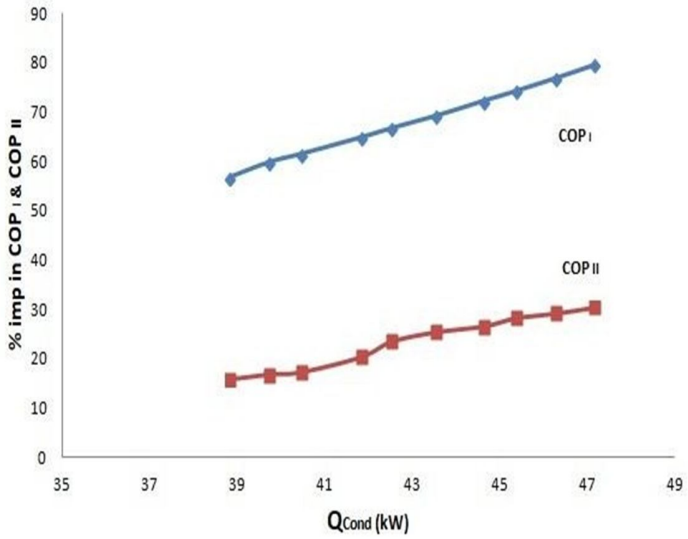


Fig.9: Single Effect

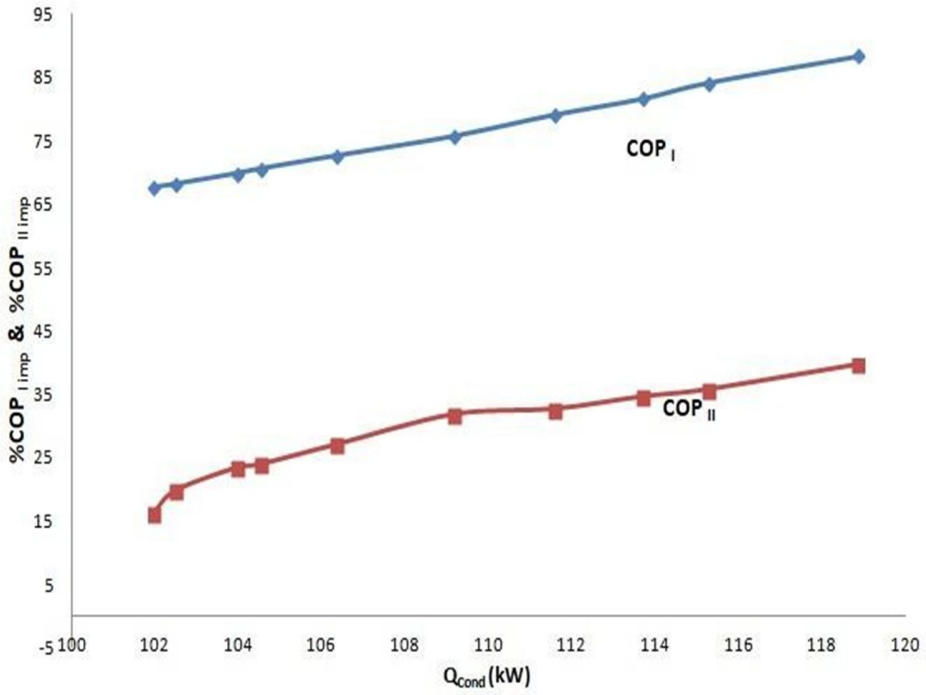


Fig.10 Double Effect

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

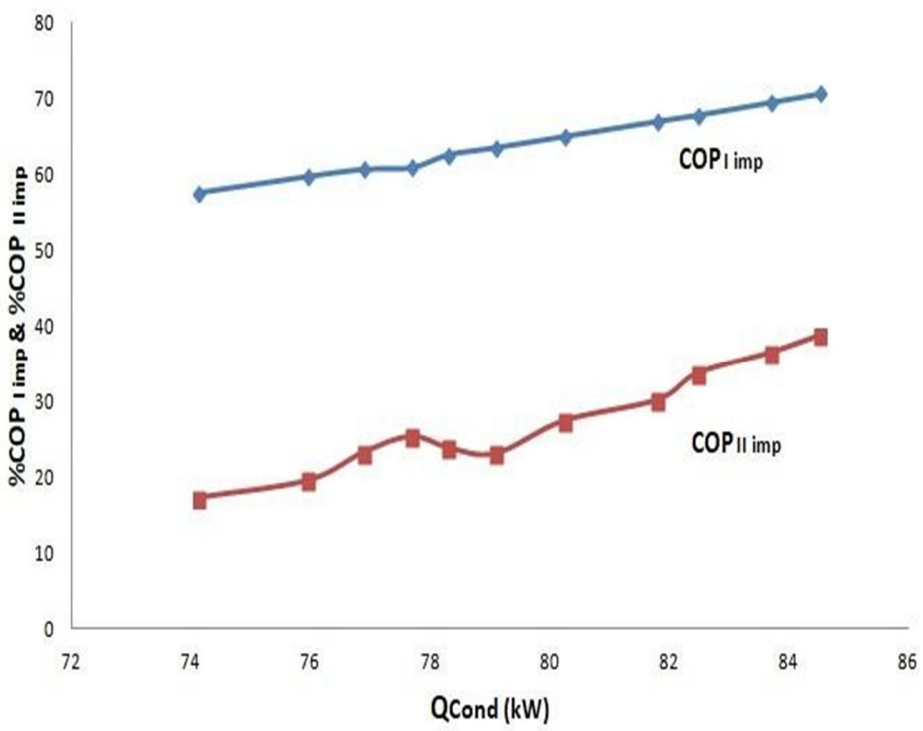


Fig.11: Half Effect

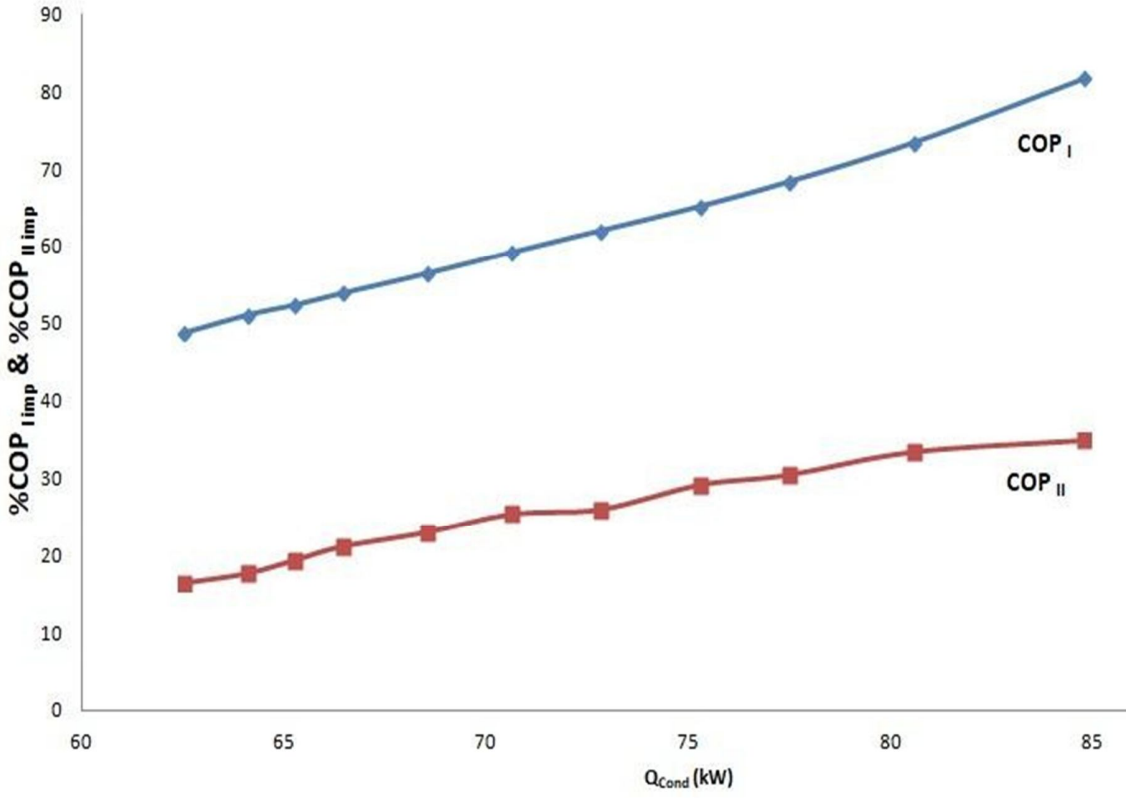


Fig.12: Triple Effect

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

C. Comparison of Percentage Improvements in COP_I & COP_{II} with Varying Temperature in Condenser of the Loop Heat Pipe (LHP):

This section shows the results on performance parameters with varying T_C for the systems. The percentage increases in the figure from Fig 13 to Fig 15 for double, half and triple effect respectively.

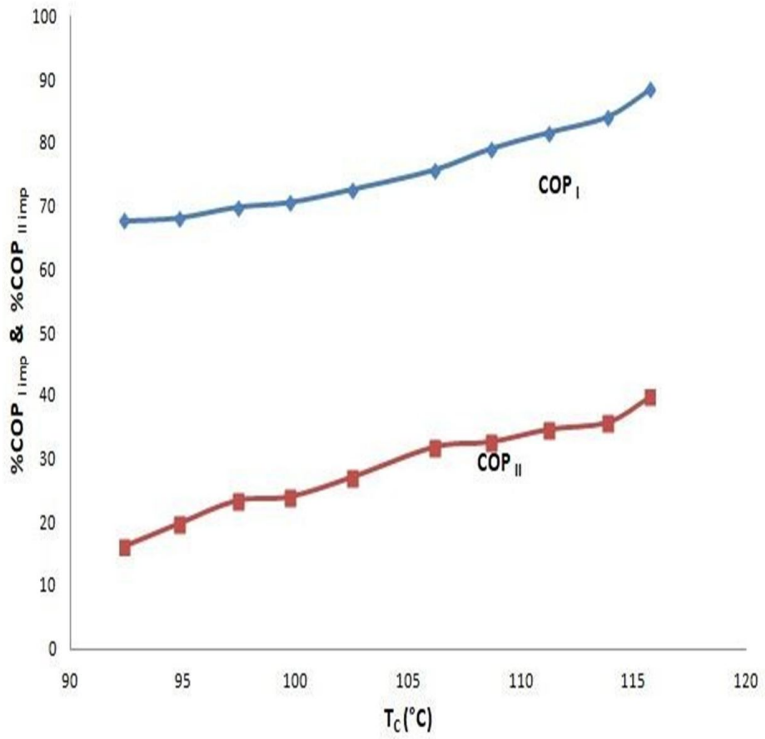


Fig.13: Double Effect

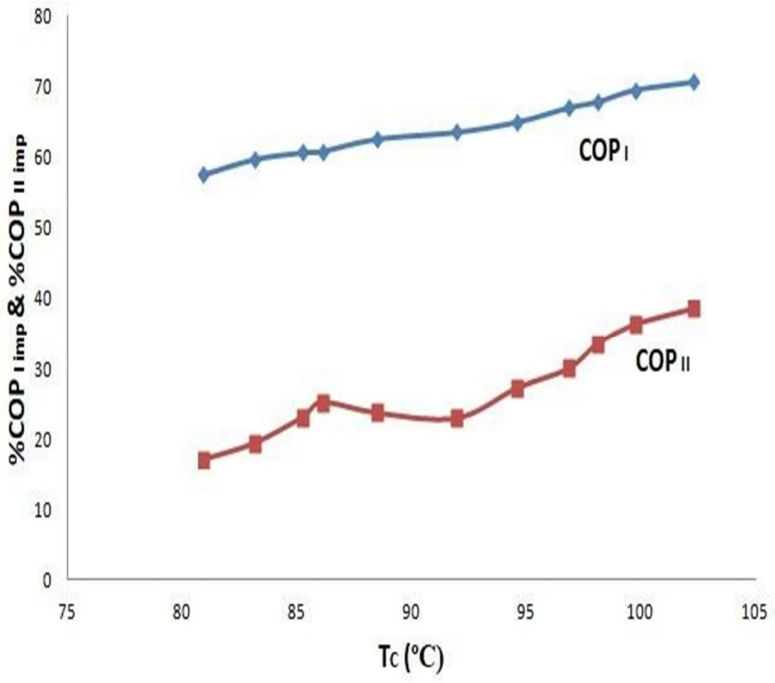


Fig.14: Half Effect

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

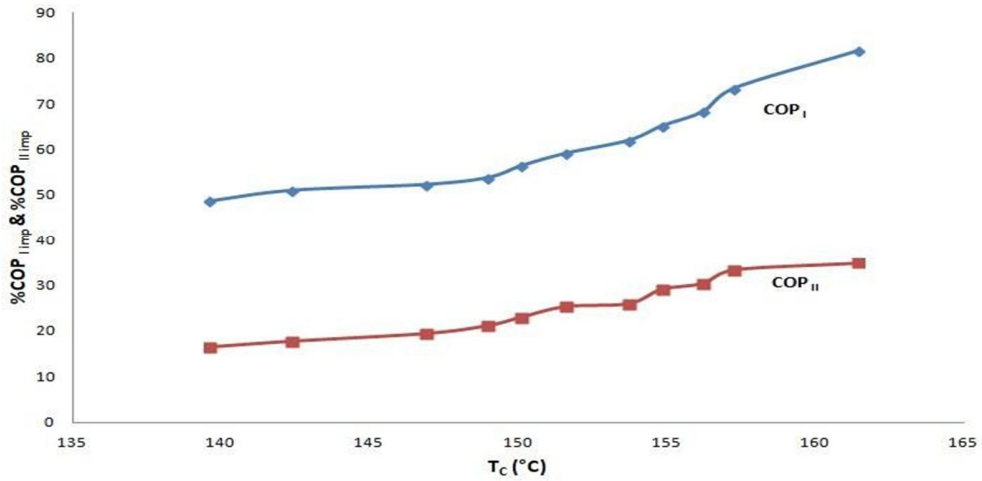


Fig.15: Triple Effect

D. Comparison of Percentage Improvements in COP_I & COP_{II} with Varying Generator Temperature:

The variation with the T_G can be seen from Fig 16 to Fig 18 for double, half and triple effect respectively.

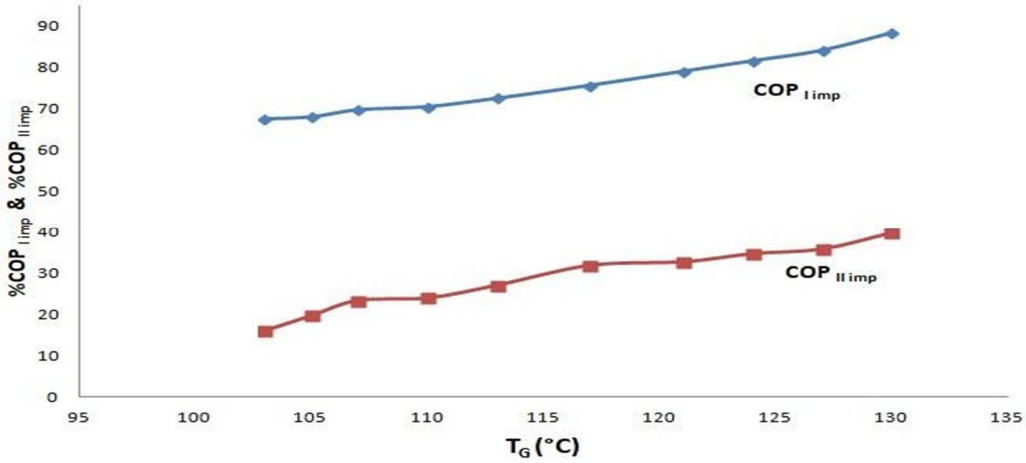


Fig.16: Double Effect

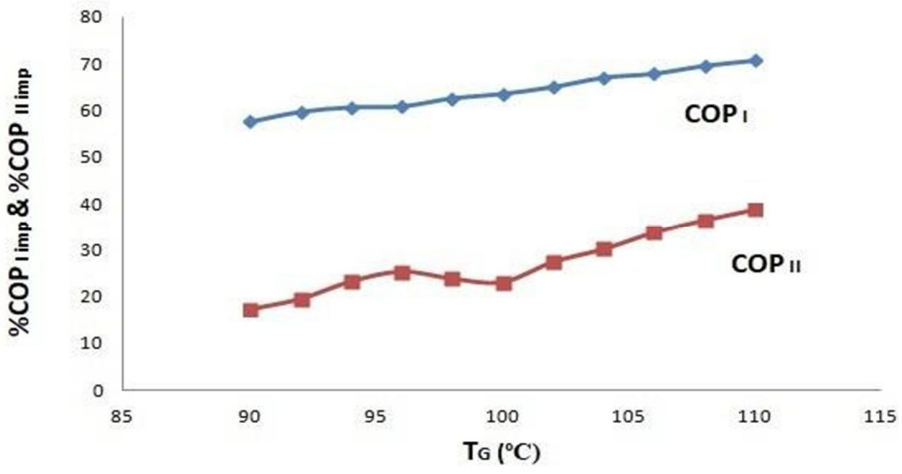


Fig.17: Half Effect

©IJRASET: All Rights are Reserved

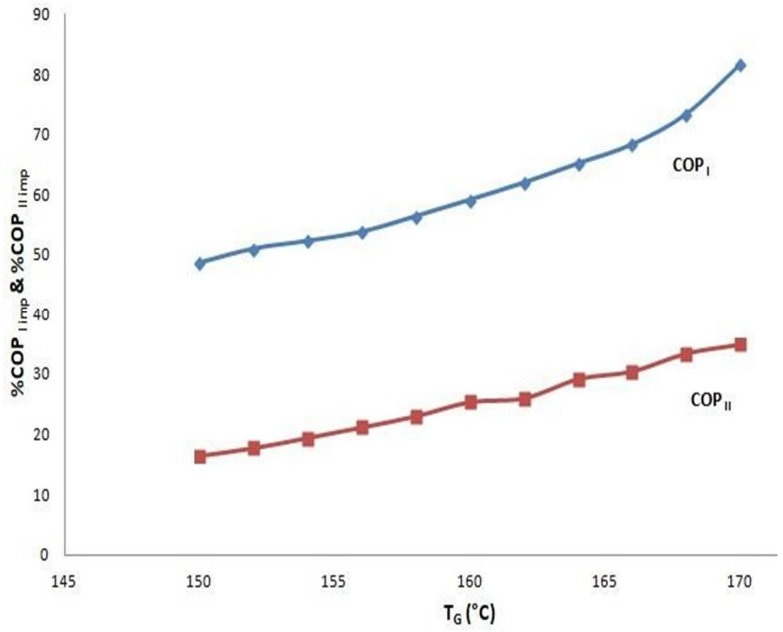


Fig.18: Triple Effect

V. CONCLUSIONS

Studies from the simulations have been presented in the previous sections. Following conclusions can be made from the results.

A. The intra cycle utilization of energy increases with the change in design, materials of the LHP and operating temperatures.

B. Maximum COP_I can be observed in Triple effect system whereas the maximum percentage increase is observed in double effect system as 78% from the initial value.

C. COP_{II} is highest for triple effect system while the increase is found to be highest for double effect system to be 28%.

D. The size of the system is gets reduced and the parts having exergy loss are eliminated along with the enhanced flexibility of the system.

E. For the systems the COP_{II} has a steeper rise over the range when compared to the COP_I for the range of heat exchange in the condenser of LHP.

F. With increase in T_G both COP_I and $COP_{II}$are increasing along with the T_C sharply.

REFERENCES

[1] Korn, F., “Heat Pipes and its Applications” Project Report 2008 MVK160 Heat and Mass Transport May 07, 2008, Lund, Sweden.

[2] T.X. Li, R.Z. Wang , L.W. Wang, Z.S. Lu, C.J. Chen,” Performance study of a high efficient multifunction heat pipe type adsorption ice making system with novel mass and heat recovery processes” International Journal of Thermal Sciences Vol.46 (2007) 1267– 1274

[3] Rajashree, R., Rao, K.S., “A Numerical Study of the Performance of Heat Pipe” Indian Journal of Pure and Applied Mathematics, 21 (1): 95-108, January 1990.

[4] M.M. Talbi, B. Agnew,” Exergy analysis: an absorption refrigerator using lithium bromide and water as the working Fuids” Applied Thermal Engineering Vol. 20 (2000) 619-630.

[5] E. Kurem,” A comparison between Ammonia-water and Water-Lithium Bromide solutions in vapour absorption heat transformers “In. Comm. Heat Mass Transfer; Vol. 28, No. 3, pp. 421-438, 2001.

[6] S.A. Adewusi, Syed M. Zubair,” Second law based thermodynamic analysis of ammonia–water absorption systems” Energy Conversion and Management Vol. 45 (2004) 2355–2369.

[7] S. Arivazhagan , R. Saravanan , S. Renganarayanan ,” Experimental studies on HFC based two-stage half effect vapour absorption cooling system” Applied Thermal Engineering Vol. 26 (2006) 1455–1462.

[8] Rabah Gomri,” Second law comparison of single effect and double effect vapour absorption refrigeration systems” Energy Conversion and Management Vol. 50 (2009) 1279–1287.

[9] S.C. Kaushika, Akhilesh Arora,” Energy and exergy analysis of single effect and series flow double effect water–lithium bromide absorption refrigeration systems” int ernational journal of refrigeration Vol.32 (2009) 1247 – 1258.

©IJRASET: All Rights are Reserved

[10] Gulshan Sachdeva, Ram Bilash," Thermodynamic Analysis of a Vapor Absorption System Using Modified Gouy-Stodola Equation" International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:8, No:12, 2014.
[11] Ankit Dwivedi, R. S. Mishra, Manjunath K," Optimization of Vapour Absorption System Using Heat Pipes" ISSN: 2321-9653 International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 5 Issue VIII, August 2017,634- 639.
[12] Ankit Dwivedi, R. S. Mishra," Performance Enhancement of Simple Vapour Absorption System Using Loop Heat Pipes" ISSN: 2321-9653 International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 6 Issue I, January 2018, 865-873.
[13] Ankit Dwivedi, R. S. Mishra," Study of Performance Enhancement of Half Effect Vapor Absorption System Using Loop Heat Pipes" ISSN: 2321-9653 International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 6 Issue I, January 2018, 1147-1156.
[14] Ankit Dwivedi, R. S. Mishra," Performance Enhancement Analysis of Double Effect Vapor Absorption System Using Loop Heat Pipes" ISSN: 2321-9653 International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 6 Issue I, January 2018,1523-1533.
[15] A Dwivedi, Thermodynamic Analysis of Vapour Compression Refrigeration System Using a Heat Pipe, Master's Degree in Delhi Technological University, 2015, 20-48
[16] Ankit Dwivedi, R. S. Mishra, Manjunath K,"Optimization of Vapour Absorption System Using Heat Pipes" ISSN: 2321-9653 International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 5 Issue VIII, August 2017,634-639.
[17] AnkitDwivedi, R. S. Mishra," Study of Performance Enhancement of Half Effect Vapor Absorption System Using Loop Heat Pipes" ISSN: 2321-9653 International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 6 Issue I, January 2018, 1147-1156.
[18] Ankit Dwivedi, R. S. Mishra," Performance Enhancement Analysis of Double Effect Vapor Absorption System Using Loop Heat Pipes" ISSN: 2321-9653 International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 6 Issue I, January 2018,1523-1533.
[19] Ankit Dwivedi, R. S. Mishra," Study of Improvement in the Performance of Triple Effect Vapor Absorption System Using Loop Heat Pipes" ISSN: 2321-9653 International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 6 Issue I, January 2018, 3160-3172.

©IJRASET: All Rights are Reserved

View publication stats

Silicon
https://doi.org/10.1007/s12633-020-00394-5

ORIGINAL PAPER

Conducting Polymer Based Gas Sensor Using PNIN- Gate All Around - Tunnel FET

Jaya Madan [1] · Rahul Pandey [1] · Rishu Chaujar [2]

Received: 14 October 2019 / Accepted: 10 January 2020
© Springer Nature B.V. 2020

Abstract
In the present work, the n + source pocket PIN gate all around tunnel FET (PNIN-GAA-TFET) based gas sensor has been proposed. Various analyte gases such as hexane, methanol, isopropanol, dichloromethane, and chloroform have been examined for their sensitivity and stability. The sensing of various analyte gases relies on the fact of work function modulation of the conducting polymer (used as gate) on the exposure of the gas vapors. Subsequently, the work function modulation of the conducting polymer gate alters the characteristics of the device that is calibrated for examining the sensitivity of the proposed gas sensor. The shift in the subthreshold current after the exposure of the analyte gas has been used as a sensing metric. Moreover, the influence of ambient temperature, conduction path, i.e., the channel radius and the oxide thickness of gate on the sensitivity and the characteristics of the proposed gas sensors have also been examined. The employment of the proposed PNIN-GAA-TFET based gas sensor may fulfill the perpetual demands of chemical and pharmaceutical industries, biomedical diagnostics, environmental monitoring and automotive industries by offering a low power, highly sensitive and reliable analyte gas detection.

Keywords Conducting polymer · Gas sensor · Sensitivity · Work function modulation · n + source pocket · Tunnel FET

1 Introduction

Employment of gas sensors in various disease diagnostics, environmental monitoring and automotive industries, to meet the demands of widespread detection and continuous monitoring, necessitate cheap, highly sensitive, stable and fast gas sensors [1, 2]. Direct transduction, CMOS compatibility, low cost, high sensitivity, low power and small size offered by the FET based gas sensor have opened up avenues in various scientific fields [3–5]. Various toxic indoor and industrial air pollutants have significant negative effects on human health, while they are also used in numerous applications [6]. Such as methanol, an organic solvent, has a wide range of applications; for the manufacturing of various colors, dyes, perfumes and drugs. However, methanol is highly toxic and fatal to human beings even at modest concentrations (<200 ppm) [7]. Moreover, the large scale usage of methanol has environmental implications [8]. Therefore, it is a requisite to sense methanol. In addition, chlorinated hydrocarbons such as dichloromethane and chloroform pose a lethal threat to the environment. The hazardous chloroform in low concentration (in the air or in water) may damage the central nervous system, liver and kidney. The major concern with chloroform is that its solubility of 8.7 g/L in water may also result in pollution [9]. It is also suspected that the chloroform is a carcinogen. It is thus necessary to sense and monitor the concentration traces of chloroform [10]. Furthermore, hexane, a toxic chemical, primarily found in common household products such as spray adhesives, stain remover and craft paints. Its frequent uses in domiciliary products often result in its exposure to humans easily [11]. Thus, hexane can be easily inhaled and absorbed through the skin, due to the easy evaporation characteristics. The consequence of inhalation of this gas may be long-lasting causing permanent nerve damage [12]. Thus, the hazardous nature of hexane necessitates its sensing and monitoring in air. Another vital volatile organic pollutant that is also widely used as a solvent and is also used as a chemical in various electronics,

✉ Rahul Pandey
rahul.pandey@chitkara.edu.in

✉ Rishu Chaujar
chaujar.rishu@dtu.ac.in

[1] VLSI Centre of Excellence, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India

[2] Microelectronics Research Lab, Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi 110042, India

Published online: 20 January 2020

pharmaceutical and other fields is isopropanol. Even at high temperatures it decomposes to toxic gas and its exposure may result in serious health issues. Thus, sensing isopropanol is always a challenge, that demands highly sensitive sensors to monitor its concentration [13]. Additionally, the environmental considerations that have arisen from flagrant misuse of natural resources by humankind over the last two centuries have, today, forced us into taking preventive measures against further environmental degradation and destruction. This serious demand for sensing various gas vapors is the motivation behind this work.

Researchers have used solid-state devices (MOSFET) as a gas sensor in which catalytic metals or conducting polymers (CP) are used as gate electrode [3, 14–17]. Designing gas sensors by utilizing a CP as a sensing layer offers high flexibility [18]. The deep ingression of analyte gases into the interior of the CP is offered by their porosity. These ingresses gas molecules lead to profound changes in the electrical properties of CP, and makes it potentially suitable for gas sensing. Another merit of employing a CP in gas sensor, is its work function tunability that aids in designing label-free gas sensors for sensing various gases. An additional merit of CP is delivered by the rapid reversible changes in the work function of CP proportional to the concentration of the gas vapors [19]. These features allow designing a gas sensor with rapid adsorption and desorption kinetics.

Further, Janata et al. [20], reports the chemical modulation of conducting polymer work function in response to the exposure of gas vapors without affecting the conductivity of the polymers. This led to the gas sensing potential of CP, and motivates the researchers to utilize the CP in FETs as a sensing layer. Additionally, Tamersit et al. [17] and Sarkar et al. [16] have reported that the sensitivity of a FET based sensor is substantially large at low subthreshold swing (SS). However, conventional FET devices show remarkable low sensitivity owing to their large SS [16, 17]. Thus, a FET with a low SS is required for realizing a gas sensor. However, as MOSFET possess a fundamental limit of 60 mV/decade on the SS. And hence TFET based gas sensors are being introduced [15] that offers high sensitivity by overcoming the limitation on SS of counterpart i.e., MOSFET. Thus, in this work, the integrated advantages of CP (work function modulation) and TFET (low SS) has been employed for designing of low cost and low power gas sensor.

In this respect, an electrochemically deposited CP as a gate in TFET is explored. The key basis of the detection of organic vapors is the work function modulation of the CP on exposure to organic gas vapors [21]. On exposure and thereby adsorption of various analyte gases, the properties of CP change and these changes modulate their work function [22]; thereby altering the characteristics of the device. Thus, the altered characteristics of the device before and after the exposure of the analyte gases allows for designing a TFET based gas sensor.

Owing to the advantages of TFET and gas sensing potential of CP, a comprehensive study of CP TFET gas sensor, for sensing gas vapors that are vital in a plethora of industries such as hexane, methanol, isopropanol, dichloromethane, and chloroform is reported in current work. This work delivers the design of cheap and reliable chemical gas sensors, which will be a huge boon to mankind as far as safe chemical handling is concerned.

2 Structural Descriptions and Simulation Methodology of Tunnel FET Based Gas Sensor

Figure 1 shows the simulation design of the n + source pocket PIN gate all around tunnel FET i.e., PNIN-GAA-TFET based gas sensor. The sensing element i.e., gate is comprised of conducting polymer with initial work function in nitrogen gas is 60 meV measured versus gold reference [23]. The working principle of the device is based on the change in work function (WF) of the CP on the exposure of gas vapors [21]. Fundamentally, by bubbling nitrogen gas through the liquid phase of the desired organic compound, the organic vapors of various gases can be obtained. The changes in the work function of CP on the exposure of gases are extracted from the earlier reported experimental data and are tabulated in Table 1 [20–23].

In addition to the CP based gate, the proposed gas sensor utilizes an n^+ source pocket (between the source and channel), that can be fabricated by epitaxial growth, to enhance the drive current capacity of the sensor. Optimized values of pocket doping and width have been accounted as obtained in our previous work i.e., pocket doping is 4×10^{19} cm^{-3} and width is 4 nm [24, 25]. The conducting path of the gas sensor i.e., the channel radius is 10 nm (default value). Before exploring the potential of PNIN-GAA-TFET as a gas sensor, the sensitivity of GAA-TFET, conventional TFET and conventional MOSFET have also been investigated for comparative analysis. For conventional TFET and GAA-TFET based gas sensor, there is no n + pocket, i.e., the source is directly followed by the intrinsic channel. The channel length is 50 nm and the gate oxide is of SiO_2 with a default thickness of 2 nm for all the devices i.e. conventional MOSFET, conventional TFET, GAA-TFET and PNIN-GAA-TFET. The source and drain doping are 1×10^{20} cm^{-3} and 5×10^{18} cm^{-3} for conventional TFET, GAA-TFET and PNIN-GAA-TFET. Further, the conventional MOSFET has a source (n^+) and drain (n^+) doping of 1×10^{19} cm^{-3}. The gas sensors are analyzed using the Silvaco Atlas device simulator [26]. The simulation setup for the TFET based gas sensor is calibrated with the already published experimental work reporting 70 mV/decade SS [27]. The calibration is done by tuning the parameter of the nonlocal band to band tunneling model i.e. the electron and hole tunneling masses from their ideal value to acquire a best-fitted

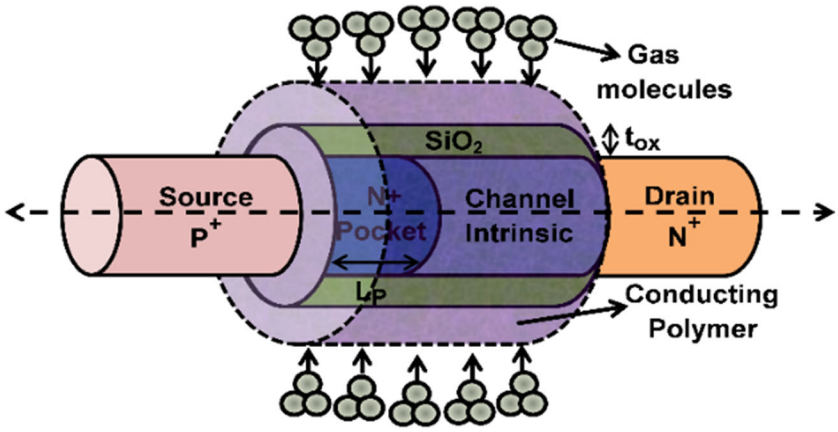


Fig. 1 Schematic diagram of the PNIN-GAA-TFET based gas sensor

I_{ds}- V_{gs} characteristics with [27] as shown in Fig. 2 that achieves SS of 68 mV/decade. In addition, the bandgap narrowing model and the Fermi Dirac statistics have been incorporated due to heavy doping of source and drain. Moreover, for incorporating the electron mobility models, concentration and field-dependent mobility models are activated. To incorporate the recombination of majority and minority charge carriers the Shockley-Read-Hall (SRH) recombination model is invoked. After the tunneling at the source-channel (for GAA-TFET and conventional TFET) or source-pocket junction (for PNIN-GAA-TFET) occurs, the charge carrier follows the conventional drift and diffusion mechanism and results in drive current, thus to consider this, the drift-diffusion model is also invoked. For the conventional MOSFET gas sensor, the concentration and field-dependent mobility model, SRH carrier recombination, Lombardi CVT mobility model and the Boltzmann transport statistics has been invoked.

3 Sensitivity Analysis of PNIN-GAA-TFET Gas Sensor for Various Analyte Gases

The impact of work function modulation (after the exposure of gas vapors) on the energy band and carrier concentration at $V_{gs} = V_{ds} = 1.0$ has been analyzed as reported in Fig. 3a–c. In channel region, results evidently shows that the difference between conduction band energy and electron quasi fermi energy is slightly higher for $CHCl_3$ and CH_2Cl_2 as compared to other three gases such as C_6H_{14}, CH_3OH and Iso-C_3H_7OH. This signifies the reduction in electron concentration for the case of $CHCl_3$ and CH_2Cl_2 in the channel region compared to other gases as depicted in Fig. 3c. Figure 3b shows the magnified view of energy bands with the electron quasi-Fermi level (EQF) and the conduction band. This clearly shows the larger difference between the conduction band and EQF for $CHCl_3$ w.r.t the case when the sensor is no exposed to gas vapors. The above situation is attributed to the elevated work function of gate electrode for CH_2Cl_2 and $CHCl_3$. Moreover, the reduced electron concentration in the channel leads to the reduction in drain current and enhancement in threshold voltage.

Figure 4 depicts the transfer characteristics of PNIN-GAA-TFET based gas sensor under the exposure of different gases and also when no gas vapors are present. In this work, the sensitivity of the gas sensor is examined as the ratio of drain current after and before the exposure of analyte gas in the subthreshold region.

$$S = I_{DS,A}/I_{DS,B}$$

Where $I_{DS,A}$ and $I_{DS,B}$ are the drain current after the sensor is exposed to analyte gas and before the sensor is exposed to the gas respectively.

It is evident from the curve that the changes in drain current under different gas analytes are remarkable large in the subthreshold region in comparison with the superthreshold regime. These distinguishable differences in drain current for

Table 1 Work function response (meV) of conducting polymer measured w.r.t Au on the exposure of various analyte gases [20, 23]

	WF_i (in N_2) w.r.t. Au	WF in gas	ΔWF response in gas
Hexane	60	80	20
Methanol	60	100	40
Iso-Propanol	60	170	110
Dichloromethane	60	210	150
Chloroform	60	215	155

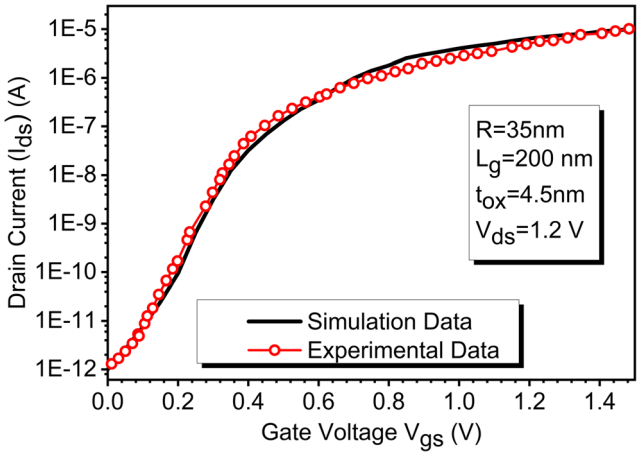


Fig. 2 Validation of simulation model with experimental results [27] for a GAA-TFET

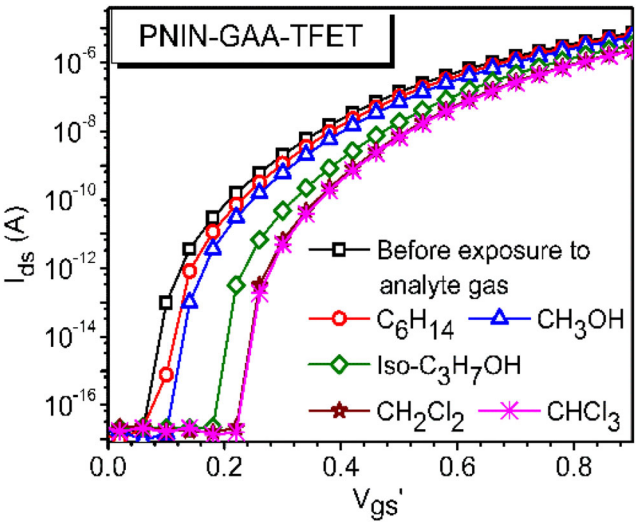


Fig. 4 Transfer Characteristics of PNIN-GAA-TFET gas sensor under different gas vapors at V_{ds} = 1.0 V w.r.t V_{gs}' = V_{gs}-0.7 V

different gases are used to calibrate the sensitivity of the device.

In Fig. 5, the sensitivity of various analyte gases for conventional MOSFET, conventional TFET, GAA-TFET, and PNIN-GAA-TFET has been presented. It is already reported in various research articles that the sensitivity is maximum in the subthreshold region and thus the sensitivity is examined at the subthreshold regime for all the devices [15, 17]. It is analyzed that for all the devices, sensitivity of hexane is least among all the analyte gases under consideration, followed by methanol, isopropanol, dichloromethane, and chloroform. This variation in sensitivity for various analyte gases is attributed to the difference in work function caused by the presence of various analyte vapors. The presence of hexane changes the work function by 20 meV whereas it is shifted by 155 meV for chloroform [20, 21, 23]. Thus, this small/large shift in work function of the conducting polymer for hexane/chloroform results in their least/highest sensitivity.

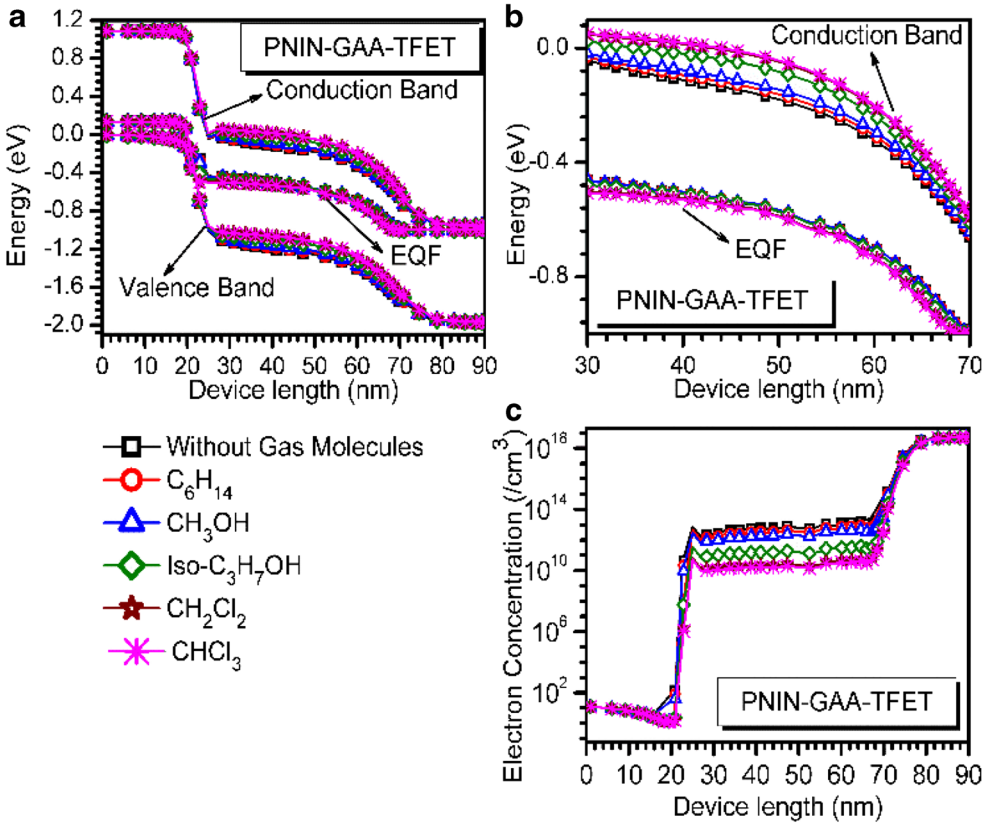


Fig. 3 **a** Energy bands of PNIN-GAA-TFET gas sensor along the device length **b** magnified view of the energy bands in the channel region **c** electron concentration of PNIN-GAA-TFET based gas sensor along the device length

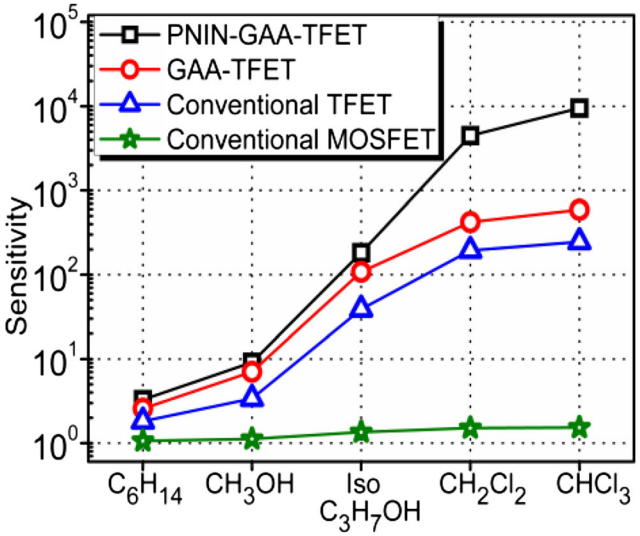


Fig. 5 Sensitivity assessment of conventional MOSFET, conventional TFET, GAA-TFET and PNIN-GAA-TFET for various analyte gases

Besides, the sensitivity of conventional MOSFET based gas sensor is least among all the gas sensors, followed by the conventional TFET, GAA-TFET, and PNIN-GAA-TFET based gas sensor. It is examined that the sensitivity of either of the TFET based gas sensor is higher as compared to the conventional MOSFET for every analyte gas. This higher sensitivity of TFETs is offered by the steeper ON/OFF transitions or the steeper SS offered by the quantum mechanical band to band tunneling that is the fundamental conduction mechanism of TFETs. Larger tunneling barrier width for the lower gate bias leads to I_{OFF} of the order of 10^{-15}A offered by the lower BTBT rate of electrons. The presence of various analyte gases alters the metal work function of the conducting polymers and thereby changes the characteristics of the device. However, the changes are maximum in the subthreshold regime and for TFETs, the changes in subthreshold regions are of many orders.

This results in substantially high sensitivity for TFETs in comparison to MOSFET. Figure 5 also shows that among TFET based gas sensor, the sensitivity is least in conventional TFET followed by GAA-TFET and PNIN-GAA-TFET. The higher sensitivity of GAA-TFET is attributed to its higher gate control capability than conventional TFET and the increased surface area of conducting polymer that allows more analyte gas reactions [15]. The integration of an n + source pocket offers steeper band bending at the source pocket junction and thereby results in remarkable alterations in the device characteristics when exposed to the analyte gases. It is evaluated that for methanol, the sensitivity of PNIN-GAA-TFET based gas sensor is 8.20 times higher than conventional MOSFET. For chloroform, the sensitivity of PNIN-GAA-TFET based gas sensor is 16.20 times, 38.79 times higher than GAA-TFET and conventional TFET respectively. Since, the sensitivity of PNIN-GAA-TFET based gas sensor is impressively high than the other devices based gas sensor under comparison, the remaining study is done only for the PNIN-GAA-TFET.

The reliance of properties of the semiconductor on temperature is the key motivation for analyzing the sensitivity of the proposed gas sensor i.e. PNIN- GAA-TFET based gas sensor at various ambient temperatures. For the good stability of the sensor, it's sensitivity should be least affected by the variations of the ambient temperature. The wide range temperature affectability (200 K–400 K) on the sensitivity of the PNIN-GAA-TFET gas sensor for various analyte gases is shown in Fig. 6. While examining the sensitivity at different ambient temperatures, it is analyzed that the sensitivity of the proposed gas sensor is high at lower temperatures.This reduction in sensitivity at higher temperatures is caused by the reduced coverage of analyte gases at high temperatures. Endurable stability, is however, attained for the wide range of temperatures. Quantitatively, it is examined that with a rise in ambient temperature from 200 K to 400 K, the sensitivity of the proposed gas sensor decreases by 2.24 times for methanol, decreases from an order of 10^6 to 10^4 for chloroform. The sensitivity being analyzed at the subthreshold regime is highly affected by the temperature caused by the exponential reliance of SRH on temperature. For TFETs, in subthreshold regime, the SRH phenomenon dominates over BTBT. Thus, a tradeoff is needed between the sensitivity and the thermal stability of the gas sensor, as for higher gate bias, the drive current is controlled by BTBT mechanism that results in marginal increase in drive current at elevated temperatures. This evaluation of sensitivity in the superthreshold regime will reduce the sensitivity but will improve the stability over a wide temperature range. The sensitivity of the proposed gas sensor is also examined at various channel radius i.e. the conduction path of the gas sensor as shown in Fig. 7. In this regard, the sensitivity

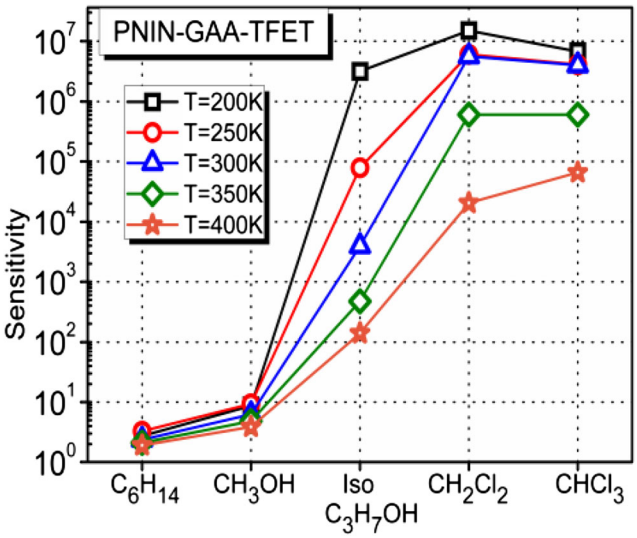


Fig. 6 Influence of temperature on sensitivity of the proposed gas sensor for various analyte gases

of the PNIN-GAA-TFET is examined at three radii i.e. 5 nm, 10 nm, 15 nm. It is evaluated that with an increase in channel conduction path or the radius of the sensor, the sensitivity of the sensor increases incredibly by many orders. The increase of sensitivity with radius is attributed to the increase in the effective surface area with radius that consequently enhances the probability of the surface reactions of the analyte gases with the conducting polymers.

The gate electrostatics is dramatically controlled by the gate oxide thickness (T_{ox}). Thus, to evaluate the role of T_{ox} in the sensitivity of the proposed gas sensor, the sensitivity of various analyte gases has been examined at three different T_{ox} i.e. 2 nm, 3 nm and 4 nm as shown in Fig. 8. As the thinner dielectric increases the effective electric field in the channel region, thus results in superior variation in the band to band tunneling. However, a thick dielectric drops large potential on itself and reduces the electric field at the tunneling junction thereby reducing the drive current [28]. The reduced barrier tunneling at lower T_{ox} yields in strong increase in the amount of sensitivity of the proposed gas sensor. Thus, the maximum sensitivity has been analyzed at $T_{ox} = 2$ nm.

After analyzing the sensitivity of the proposed gas sensor, its electrical performance is also examined in terms of the peak of the band to band tunneling rate of electrons at the junction of source and n + pocket and the threshold voltage of the sensor. Figure 9a reflects the influence of temperature variation on the BTBT rate of electrons on the various analyte gases of the proposed gas sensor. The considerable enhancement in the BTBT rate of electrons at elevated temperature is evidently shown in Fig. 9a. The bandgap narrowing offered by the rise in ambient temperature results in the narrowing of the tunneling barrier width and thereby enhances the BTBT rate of electrons. Thus, for each analyte gas vapor, the increase in temperature increases the BTBT rate of electrons. Moreover, the

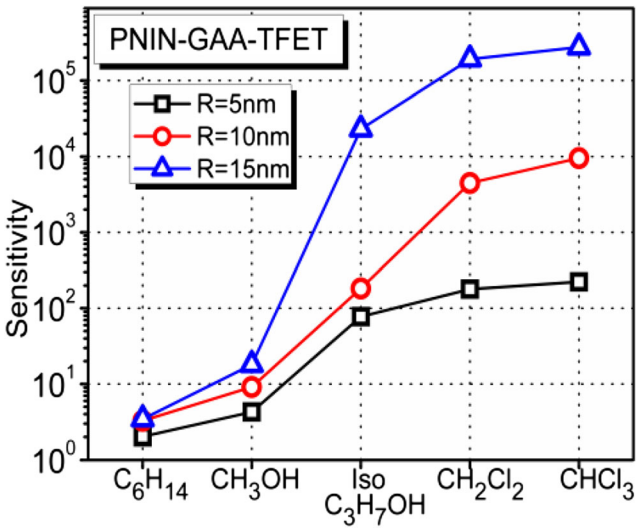


Fig. 7 Influence of channel radius on sensitivity of proposed gas sensor for various analyte gases

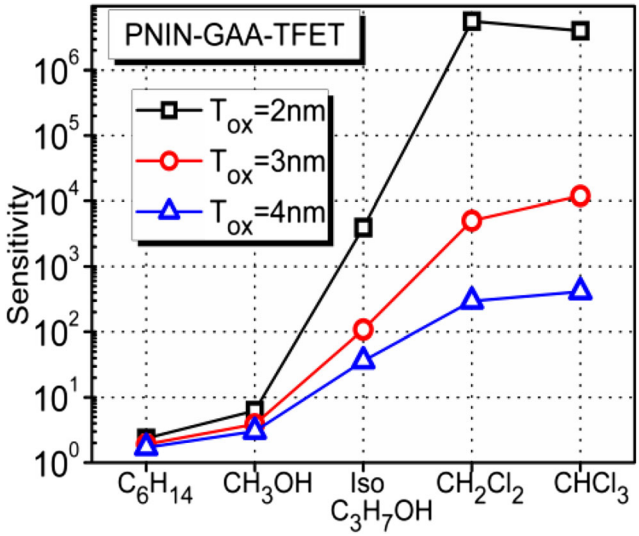


Fig. 8 Influence of gate oxide thickness on sensitivity of proposed gas sensor for various analyte gases

BTBT rate of electrons is highest/lowest for chloroform/hexane. The enhancement in BTBT rate of electrons at elevated temperature yields early switching of the sensor as clearly depicted in Fig. 9b. Furthermore when there is no gas molecules exposure, the threshold voltage of the sensor (defined as $V_{gs}@I_{ds} = 10^{-7}$A) is minimum. However, when the sensor is exposed to the analyte gases, the workfunction modulation of the conducting polymers reduces the BTBT rate of electrons and thereby increases the V_{th} of the proposed gas sensor as illustrated in Fig. 9b.

Figure 10a depicts the influence of the proposed gas sensor radius on the peak of BTBT rate of electrons at the junction of source and n + pocket for various analyte gas vapors. As reported earlier by [29, 30], for smaller radius, the electron mobility decreases but the gate controllability over the device is enhanced that enhances the lateral electric field at the tunneling junction, which enhances the band bending and thus, reduces the barrier width for tunneling of electrons and henceforth increases the BTBT rate of electrons substantially, as depicted in Fig. 10a. With an increase in the channel conduction path i.e. the channel radius, V_{th} of the proposed sensor increases i.e. the sensor switches ON at higher gate bias as illustrated in Fig. 10b. Additionally, the reduced BTBT rate of electrons for hexane to chloroform increases the V_{th} of the sensor. Further, the impact of gate oxide scaling on the non-local BTBT rate of electrons of the proposed gas sensor for the presence of various analyte gas vapors has also been studied and is shown in Fig. 11a. It is clearly depicted that for thicker/thinner gate oxide, the BTBT rate of electrons is minimum/maximum. The enhancement of BTBT at thinner gate oxide is due to the enhanced electric field that upshots in the higher probability of tunneling rate. Furthermore, it is examined that the BTBT rate of electrons decreases by 16 times with an increase in T_{ox} from 2 nm to 4 nm and it reduces from an

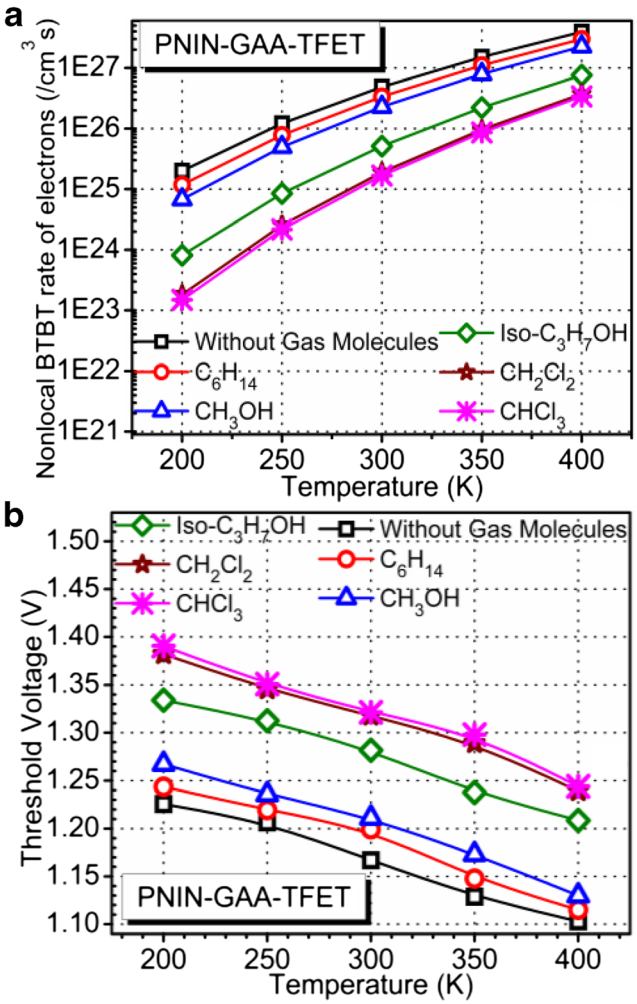


Fig. 9 Influence of temperature on **a** nonlocal band to band tunneling rate of electrons and **b** threshold voltage of proposed gas sensor for various analyte gases

order of 10^{25} to 10^{22} for dichloromethane and chloroform. The influence of T_{ox} on V_{th} is depicted in Fig. 11b. The reduced BTBT for thicker gate oxide leads to increase in V_{th}. It is evaluated that with an increase in T_{ox} from 2 nm to 4 nm, V_{th} of the sensor increases by 15% and 13% for the case when no gas is present and for chloroform respectively.

4 Conclusion

This work presents a comprehensive comparative analysis of the operation of major candidates for FET based sensors viz. MOSFETs, conventional TFETs, GAA-TFETs and PNIN-GAA-TFETs. This analysis clearly indicates the superiority of the PNIN-GAA-TFET in terms of both, sensitivity and stability. Such superiority is attributed to the higher gate controllability and steeper band bending in these devices. Therefore, the PNIN-GAA-TFET forms the basis of the work done in the rest of the analysis. The sensing action of the proposed sensor is based on a relatively well known physical mechanism. Earlier chemical analyses have revealed and extensively studied the adsorption of organic gases on conducting polymer. Using this principle as the primary basis of operation, a gas sensor with conducting polymer forming the gate electrode is proposed. The presented work further focuses on optimizing the operation of the sensor w.r.t its parameters such as channel radius and oxide layer thickness. It is found that, for the device under consideration, maximum sensitivity is obtained for a channel radius of 15 nm and a gate oxide thickness of 2 nm. Optimization of this sort is essential for any commercially viable sensor. Furthermore, the temperature affectability of the sensor is also considered with regard to exploring the stable range of operation of the sensor. This analysis reveals that the proposed sensor has a stable range of 200 K over 200 K- 400 K, with highest sensitivity at 300 K. This temperature stability is appropriate for most applications indicating the industrial feasibility of PNIN-GAA-TFET based sensors. Given the results obtained here, one can be

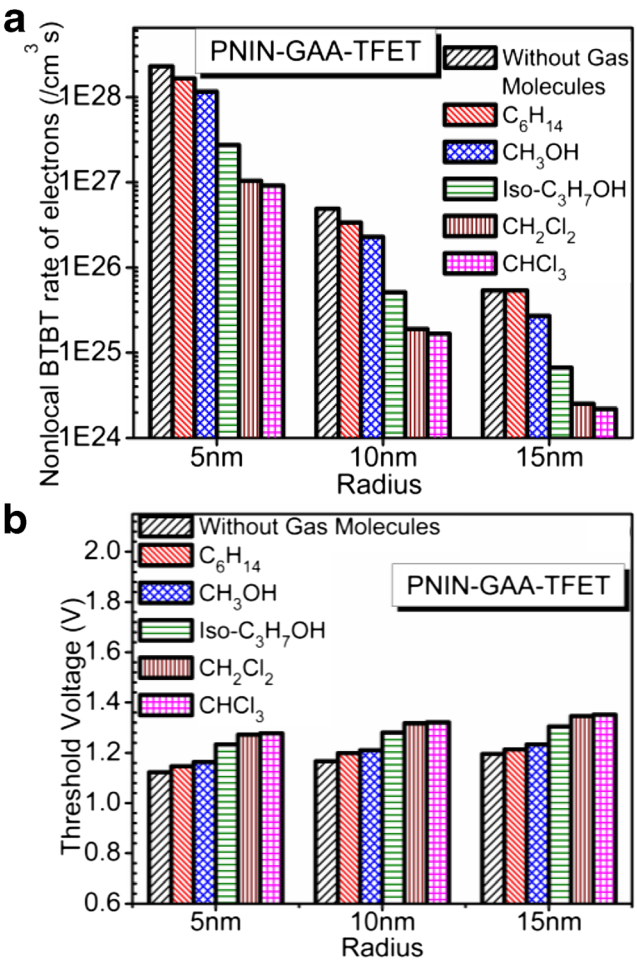


Fig. 10 Influence of channel radius on **a** nonlocal band to band tunneling rate of electrons and **b** threshold voltage of proposed gas sensor for various analyte gases

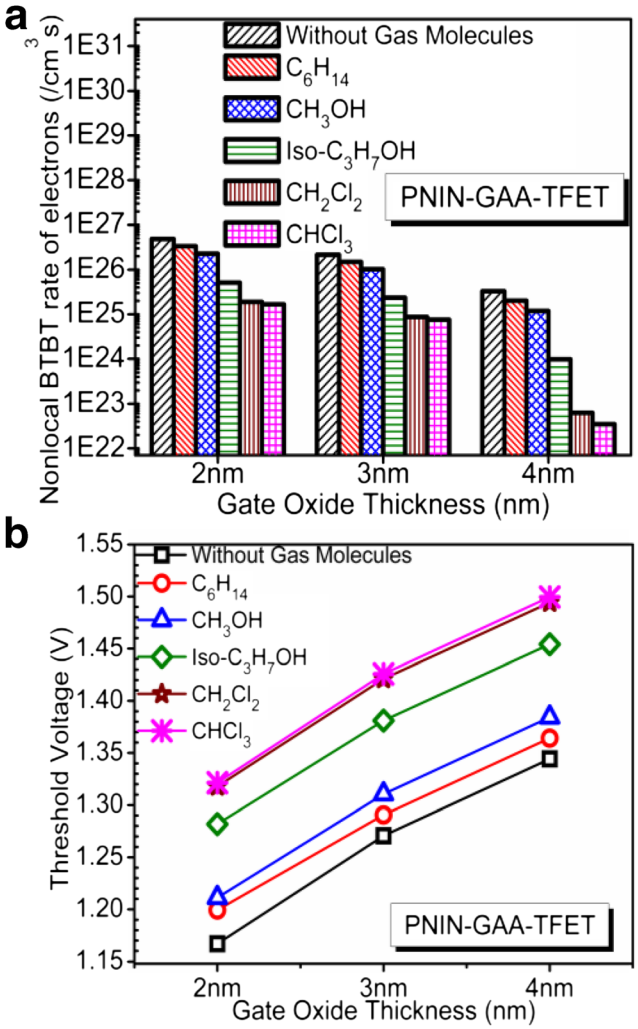


Fig. 11 Influence of gate oxide thickness on **a** nonlocal band to band tunneling rate electrons and **b** threshold voltage of proposed gas sensor for various analyte gases

confident that further research and development followed by the fabrication of the proposed PNIN-GAA-TFETs is certain to yield promising results.

Acknowledgments The authors gratefully acknowledge the support from Microelectronics Research Laboratory, Delhi Technological University to carry out the work. The author, Dr. Jaya Madan thanks to Dr. Rajnish Sharma from Chitkara University for his continuous support during the preparation of the manuscript.

References

1. Jayaswal N, Raman A, Kumar N, Singh S (2019) Design and analysis of electrostatic-charge plasma based Dopingless Igzo vertical nanowire Fet for ammonia gas sensing. Superlattice Microst 125: 256–270
2. Deng Y (2019) Sensing devices of semiconducting metal oxides gas sensors', semiconducting metal oxides for gas sensing. Springer
3. Lundström I, Armgarth M, Spetz A, Winquist F (1986) Gas sensors based on catalytic metal-gate field-effect devices. Sensors Actuators 10(3–4):399–421
4. Pourasl AH, Ariffin SHS, Ahmadi M, Gharaei N, Rashid RA, Ismail R (2019) Quantum capacitance model for Graphene Fet-based gas sensor. IEEE Sensors J 19(10):3726–3732
5. Mokkapati S, Jaiswal N, Gupta M, Kranti A (2019) Gate-all-around nanowire junctionless transistor-based hydrogen gas sensor. IEEE Sensors J 19(13):4758–4764
6. Akbari E, Nilashi M, Buntat Z (2019) Analytical investigation of superior gas sensor based on phosphorene. Microsyst Technol 25(3):897–903
7. Banerjee N, Roy S, Sarkar CK, Bhattacharyya P (2013) High dynamic range methanol sensor based on aligned Zno nanorods. IEEE Sensors J 13(5):1669–1676
8. Sahay P, Nath R (2008) Al-doped Zno thin films as methanol sensors. Sensors Actuators B Chem 134(2):654–659
9. Fong JK, Pena JK, Xue Z-L, Alam MM, Sampathkumaran U, Goswami K (2015) Optical sensors for the detection of trace chloroform. Anal Chem 87(3):1569–1574
10. Sharma S, Nirkhe C, Pethkar S, Athawale AA (2002) Chloroform vapour sensor based on copper/polyaniline nanocomposite. Sensors Actuators B Chem 85(1):131–136
11. Zheng J, Zhang W, Cao J, Su X, Li S, Hu S, Li S, Rao Z (2014) A novel and highly sensitive gaseous N-hexane sensor based on thermal desorption/cataluminescence. RSC Adv 4(41):21644–21649
12. Yang P, Ye X, Lau C, Li Z, Liu X, Lu J (2007) Design of efficient zeolite sensor materials for N-hexane. Anal Chem 79(4):1425–1432
13. Li S-H, Chu Z, Meng F-F, Luo T, Hu X-Y, Huang S-Z, Jin Z (2016) Highly sensitive gas sensor based on Sno2 nanorings for detection of isopropanol. J Alloys Compd 688:712–717
14. Josowicz M, Janata J (1986) Suspended gate field effect transistors modified with polypyrrole as alcohol sensor. Anal Chem 58(3): 514–517
15. Madan J, Chaujar R (2016) Palladium gate all around - hetero dielectric -tunnel Fet based highly sensitive hydrogen gas sensor. Superlattice Microst 100:401–408
16. Sarkar D, Gossner H, Hansch W, Banerjee K (2013) Tunnel-field-effect-transistor based gas-sensor: introducing gas detection with a quantum-mechanical transducer. Appl Phys Lett 102(2):023110
17. Tamersit K, Djeffal F (2016) Double-gate graphene nanoribbon field-effect transistor for DNA and gas sensing applications: simulation study and sensitivity analysis. IEEE Sensors J 16(11):4180–4191
18. Hangarter CM, Bangar M, Mulchandani A, Myung NV (2010) Conducting polymer nanowires for chemiresistive and Fet-based bio/chemical sensors. J Mater Chem 20(16):3131–3140
19. Persaud KC (2005) Polymers for chemical sensing. Mater Today 8(4):38–44
20. Janata J, Josowicz M (2003) Conducting polymers in electronic chemical sensors. Nat Mater 2(1):19–24
21. Janata J, Josowicz M (1998) Chemical modulation of work function as a transduction mechanism for chemical sensors. Acc Chem Res 31(5):241–248
22. Janata J, Josowicz M (1997) Peer reviewed: a fresh look at some old principles: the kelvin probe and the Nernst equation. Anal Chem 69(9):293A–296A
23. Blackwood D, Josowicz M (1991) Work function and spectroscopic studies of interactions between conducting polymers and organic vapors. J Phys Chem 95(1):493–502
24. Madan J, Chaujar R (2017) Gate drain Underlapped-Pnin-Gaa-Tfet for comprehensively upgraded analog/Rf performance. Superlattice Microst 102:17–26

25. Madan J, Chaujar R (2017) Numerical simulation of N+ source pocket pin-Gaa-tunnel Fet: impact of Interface trap charges and temperature. IEEE Trans Electron Devices 64(4):1482–1488
26. Silvaco I (2011) Atlas User's manual. Santa Clara, CA, Ver, 5
27. Chen Z, Yu H, Singh N, Shen N, Sayanthan R, Lo G, Kwong D-L (2009) Demonstration of tunneling Fets based on highly scalable vertical silicon nanowires. IEEE Electron Device Letters 30(7): 754–756
28. Knoch J, Mantl S, Appenzeller J (2007) Impact of the dimensionality on the performance of tunneling Fets: bulk versus one-dimensional devices. Solid State Electron 51(4):572–578
29. Lee JS, Choi WY, Kang IM (2012) Characteristics of gate-all-around hetero-gate-dielectric tunneling field-effect transistors. Jpn J Appl Phys 51(6S):06FE03
30. Verhulst AS, Sorée B, Leonelli D, Vandenberghe WG, Groeseneken G (2010) Modeling the single-gate, double-gate, and gate-all-around tunnel field-effect transistor. J Appl Phys 107(2):024518

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Corporate entrepreneurship in India: The road less travelled

S. S. Khanka
University School of Management & Entrepreneurship,
Delhi Technological University (Formerly Delhi College of Engineering),
&
National Trainer on Ethics and Values in Public Governance recognized by Department of Personnel and Training (DoPT), Government of India in collaboration with the United Nations Development Programme (UNDP)

Keywords
Entrepreneurship, Corporate Entrepreneurship, Intrapreneurship, Innovation, Organizational Culture, Entrepreneurship Development Programmes (EDP's

Abstract

Change in business has become sine qua non with change in business environment. Darwin's Law of the 'survival of the fittest' suggests either change or perish. In order to adapt to the changing business environment, corporations must keep on changing by thinking uniquely and entrepreneurially. Thinking entrepreneurially, or say, innovatively in the existing corporations is called 'corporate entrepreneurship'. The notion and practice of corporate entrepreneurship have become very dominant and has been on increase in the western countries. However, corporate entrepreneurship could so far not make any perceptible headway in the Indian corporations. Indian corporations have yet to trudge on this relatively less known path called 'corporate entrepreneurship.'

It is against this backdrop; an attempt has made in this paper to present an overview of this relatively less travelled path of corporate entrepreneurship in India. Accordingly, it covers the aspects like what corporate entrepreneurship is, how it is different from general entrepreneurship, what are its drivers, prerequisites and barriers, and what makes corporate entrepreneurship successful.

Introduction

Of late, a new breed of entrepreneurship has emerged known as '*corporate entrepreneurship'* also called *'intrapreneurship.'* As this type of entrepreneurship is emerging with increasing scale and size world-over, it has, therefore, been attracting increasing attention and interest of researchers in the literature of entrepreneurship. Corporate entrepreneurship is an evolutionary concept and is of relatively recent origin in India. As limited is the development of corporate entrepreneurship in India, much more limited is availability of literature / research evidence on whatsoever limited corporate entrepreneurship has developed in the country. In the absence of adequate availability of data and literature on corporate entrepreneurship, our understanding about corporate entrepreneurship in India remains, thus, far from adequate and satisfactory. As such, our understanding on corporate entrepreneurship in India is based mainly on assumptions, predilections and suppositions. It is against this backdrop; the present paper makes a modest attempt to present a systematic overview of corporate entrepreneurship in India. The paper accordingly addresses two major aspects of corporate entrepreneurship like: what is corporate entrepreneurship and how is it different from general entrepreneurship, what are its drivers, prerequisites and barriers, and what makes corporate entrepreneurship successful? These are discussed in seriatim.

Meaning of Corporate Entrepreneurship

The new breed of entrepreneurship emerging within the organizations is called *'Corporate Entrepreneurship'* or '*Intrapreneurship.*' Corporate entrepreneurship emerges from within the confines of an existing enterprise. In big organisations, the top executives are encouraged to catch hold of new

ideas and then convert these into products and services through research and development activities within the framework of organisation. The concept of intrapreneurship has become very popular in developed countries like America. It is found that an increasing number of intrapreneurs are leaving their jobs in big organisations and are starting their own enterprises. Many of such intrapreneurs have become exceedingly successful in their ventures. They are causing a threat to the organisations they left.

Different thinkers have defined corporate entrepreneurship differently. For example, Steve Jobs had defined the term corporate entrepreneurship as: "*a group of people going, in essence, back to the garage, but in a large company.*"

According to Gifford Pinchot (1985), "*Intrapreneur is an entrepreneur within an already established organization.*"

Pramodita Sharma and James J. Chrisman (1999: 11-28) have given an extensive definition of corporate entrepreneurship as "*a process whereby an individual (or a group of individuals) in association with an existing organization, creates a new organization or instigates renewal or innovation within the organization. Under this definition, strategic renewal (which is concerned with organizational renewal involving major strategic and/or structural changes, innovation (which is concerned with introducing something new to the market place), and corporate venturing (corporate entrepreneurial efforts that lead to the creation of new business organizations within the corporate organization) are all important and legitimate parts of the corporate entrepreneurship process.*"

Thus, corporate entrepreneurship is the process by which organizational members within an established company conceive, foster, launch, and manage a new form of business that is distinct in certain respects from the parent company and leverages the parent company's assets, resources, and market.

Corporate Entrepreneurship is Different from Entrepreneurship

That corporate entrepreneurship (also known as intrapreneurship) is different from entrepreneurship (Pinchot 1985) is presented in the following Table 1.

Table 1: Difference between Corporate Entrepreneurship and Entrepreneurship

Difference	Entrepreneurship	Corporate Entrepreneurship
1. Approval	Needs to gather nerve.	Must gather approval.
2. Dependency	Entrepreneur is independent in his / her business operations.	But, intrapreneur is dependent on the entrepreneur, i.e. the owner of the organization
3. Fund	Entrepreneur raises funds himself / herself.	Intrapreneur does not raise funds, but the organization wherein he / she works, raises and provides funds.
4. Search	Entrepreneur searches markets to acquire resources for new venture / enterprise.	Intrapreneur looks inside the organization itself for resources that are so far not used efficiently.
5. Risk	Entrepreneur bears the risk involved in entrepreneurship	Intrapreneur does not bear risk, but the organization bears risk.
6. Operation	Entrepreneur operates from outside	Intrapreneur operates from within the organization itself.
7. Organization	Entrepreneurship creates a new enterprise.	It brings new idea(s) to the already existing organization.
8. Ownership and Control	Entrepreneur owns and controls his / her business, Ownership and control are inseparable in case of entrepreneurship.	Intrapreneur does not own the business, only controls it. Thus, ownership and control are separable in case of corporate entrepreneurship.

Why Corporate Innovation is needed in India?

The business environment is full of uncertainties in terms of market conditions, competition level, customer choice and technological advancements. These render what was workable in the past as obsolete at present. Obviously, only the corporations that fit for the requirements of the time survive and thrive. For this, corporations have either to do something altogether new or add new value to the existing one to make it relevant to the requirement of the time. This is called innovation that enables corporations to fit for the purpose. Following are some major reasons why innovation is significant for modern day corporations in India:

1. Increasing Competition: Ever increasing competition in the market has emerged as one of the most driving forces for innovation in the business world. Organizations are bound to think and act in a newer and innovative manner to have competitive advantage to survive and thrive in today's highly competitive environment. One cannot thrive in today's competitive environment by just working in a stereotypical manner while consumers' choices and preferences towards products and services have been on continuous change. This is placing burdens on organizations to keep up, either by matching product features or attempting to leapfrog over competitors by offering altogether newer and better products and services to the consumers. While the former approach places a limit on the potential for organizational success, the latter one increases potential for success. Evidence suggests that the successful organizations do not follow others, but one who strive to be the ones to emulate. This can only be done through innovation.

2. Increasing Globalization: By now, market has been shrinking due to globalization. The global market is characterized by large number of customers with varying tastes and preferences towards products and services scattered over vast territory of the world. So, to say, organizations have now access to global market. It has emerged as an opportunity for corporations to serve global customers by identifying new markets and offering newer and better products and services to the customers locates over vast global territory. This calls for implementation of innovations by corporations to thrive in the highly competitive global market.

3. Increasing Consumer Expectations: With advancement in social media and information technology, on the one hand, and availability of increasing options for products and services, on the other, consumers have been becoming increasingly knowledgeable and choosy about what to use so as to improve their standard of living. As such, consumers are not to accept mediocre or obsolete products or services. Such market scenario has primed corporations to innovate, regardless of their industry, and make it core to their business strategy. Failing in doing so means risking becoming irrelevant to consumers.

4. Increasing Technological Advancement: In the face of continuous advancement in technology, technology which was earlier a feature of large corporations, has become now an almost must feature of new start-ups in the Make in India philosophy. Start-ups are often launching with or driving the new technologies. Technology driven innovation is becoming the norm of the day. For example, 3D Printing is hitting critical mass. Thus, using new technologies in a new way and developing new technologies through innovation has become almost essential for corporations to meet the challenge posed by increasing advancement in technology.

5. Changing Workforce Demographics: Corporations have been experiencing changing workforce demographics on the horizon for quite some time. Workforce demographics is experienced not only in terms of employees with diverse backgrounds in terms of education, knowledge, and cultures, but also in terms of generational gaps. For example, while the number of baby-boomers has been almost to retire, the number of Generation Y employees is on continuous increase. This is also posing the problem of talent acquisition and retention in already shortage of talented manpower available in the market. This necessitates corporations to become innovative in their approach towards human resource practices ranging from attraction to development to retention of employees. Of late, integration of 3Ps, i.e. Profit, People and Planet has emerged as a new challenge for corporations to survive in today's highly sensitive society. Addressing this new

challenge obviously requires innovative approaches and solutions. Corporations that follow the approach have been experiencing competitive advantage to thrive in the business environment of the time.

What are the Drivers of Corporate Entrepreneurship in India?

Corporate entrepreneurship essentially involves replacing the *status quo* by introducing change in manner of doing business. Change is generally resisted by the stakeholders. Then question arises is what drives change to replace *status quo.* Research evidence (Khanka 2009: 209-218) suggests that there are certain drivers that facilitate the emergence and introduction of entrepreneurship and for that matter corporate entrepreneurship in an existing organization. These include but are not limited to the following only:

Organization Culture: Innovative thinking as a norm of organizational culture serves as a key driver in developing entrepreneurial mind-set among the employees in the organization. More intensively the employees hold the belief in innovative thinking, greater is its effect on fostering entrepreneurial mind-set among the employees and *vice versa*.

Organizational Structure: As against rigid organizational structure, a flexible structure serves as more effective driver to effect innovative thinking among the employees. The reason is flexibility in structure provides employees required scope and freedom in doing their works.

Organization Policy: More favourable the organizational policy, more possibilities for innovative thinking for employees and vice versa. Here, 3M Co.'s policy of allowing its engineers and scientists to spend 15% of their time on projects of their own design is worth mentioning.

Top Management Support: No idea, no matter how much new and promising it is, can fructify without approval and needed support from the top management. In fact, assurance of support from top management serves as an encouragement for employees to search for new ideas and implement the same.

Resource Availability: Like any activity, innovative thinking cannot happen in vacuum but on the availability of required resources. Thus, the availability of required resources also serves as a driver to foster innovative thinking in the organization.

Risk-Taking and Failure Tolerance: That organization is ready to take risk in implementing new ideas and accept failure, if any, as a learning experience also serve as effective drivers to effect innovative thinking in the organization.

Compensation and Incentive System: Provision for compensation and incentives to the employees generating new ideas and wealth also serve as drivers to inject innovative thinking among the employees. Dr. Nelson Levy, a former Vice President of Research and Development and President of various global pharmaceutical companies once quipped, "*I might as well give my people 15% paid leave!*"

Possible Forms of Corporate Entrepreneurship

Corporate entrepreneurship can take place in the following different forms:

Entering New Business: Entering into a new business by an established organization is one of the main forms of corporate entrepreneurship. An established organization due to various reasons like decline in the existing business or identifying the availability of new emerging business opportunities in the given environment, may be interested to enter into a new business venture. Examples of existing corporations entering new business abound in corporate India. Bharati Enterprises, Tata Sons, Reliance Industries, etc. are such examples to name a few.

Championing New Product Idea: The other form of corporate entrepreneurship is a new product idea championed by an individual or individuals within a corporate context. In fact, this is the most famous form of corporate entrepreneurship when managers / executives can apply their creative thinking to evolve some new product ideas for the organization. For example, in Bharati Enterprises, professionals are required to perform in an entrepreneurial manner, called '*professional entrepreneurs,*' to find out better way to do the same function. With this professional entrepreneurial

spirit, Bharati has come up with new solutions to old problems, newer versions to existing solutions, new solutions to new problems, and even paradigm shifts in its business. This innovative thinking has made Bharati the first telecom player in India to outsource its functions such as IT network, call centres, distributions, and more recently tower companies.

Creation of Entrepreneurial Philosophy: Fostering entrepreneurial philosophy or culture within the entire organization is yet another form of corporate entrepreneurship. This is a situation when efforts are made to develop entrepreneurial mindset among the organizational members so that entrepreneurial philosophy permeates throughout the organization. For example, Google follows this philosophy by creating an environment in which all Googlers think and act in an innovative or entrepreneurial manner. This entrepreneurial philosophy has been a lynchpin behind every success of Google and other successful organizations like Bharati

Besides above factors, there are yet other conditions too necessary to make corporate entrepreneurship successful. These include being gender neutral, independent thinking, fostering a questioning culture and engendering a more egalitarian set up

Models of Corporate Entrepreneurship

Based on an in-depth research study of 30 corporations as diverse as Google, DuPont, and Cargill, Robert Wolcott and Michael Lippitz (2007) reached to the conclusion that approach to corporate entrepreneurship, depending on two dimensions under the direct control of management is likely to vary from company to company. The two dimensions are: (1) Organizational Ownership, i.e., who within the organization has primary ownership for the creation of new businesses? (2) Resource Authority i.e., is there a dedicated fund allocated to corporate entrepreneurship? Together the two dimensions generate a matrix with four dominant models of corporate entrepreneurship:

A brief description of each of these follows in seriatim.

The Opportunist Model: There are ample evidences to mention that most of companies begin with what is called the *"Opportunist"* model of corporate entrepreneurship. Such corporate entrepreneurship takes place when some intrepid project champion without any designated organizational ownership or resources proceeds to seize available opportunity. In fact, this is the typical way how a start-up company turns an intrepid visionary's dream into reality. The intrepid champions especially in large established companies, must toil against the organizational odds of one type or other. Therefore, corporate entrepreneurship to be successful within opportunist model requires favourable organizational environment when top management is supportive and open to experimentation. Without such type of supportive environment, good ideas can easily fall through organizational cracks or receive insufficient funding to prove feasible.

The Enabler Model: The enabler model of corporate entrepreneurship is based on the premise that employees in an organization will be willing to develop new ideas / concepts provided they are provided adequate and required support. Employees are not given any formal form of ownership to pursue new ideas but are provided with clarity on which opportunities to pursue, guidelines for funding, recruitment and retention of entrepreneurially minded employees and, above all, active support from top management. Google Inc. presents the excellent enabler model of corporate entrepreneurship where employees at Google can spend 20% of their time to promote their ideas to colleagues, develop their concepts and build prototypes. It is important to mention that at Google no preconceived criteria or hurdle apply to a project. The project can continue till it appears to have potential to serve the interest of Google employees.

The Advocate Model: In the advocate model, company assigns ownership to a group of organizational members for the creation of new businesses with modest provision for funds. Here is an excellent example of the advocate model of corporate entrepreneurship practiced by E.I. du Pont de Nemours and Co., the 200-year-old global conglomerate. When CEO Chad Holliday, in 1999, realized that the company's growth had declined, Clad Holliday assigned DuPont veteran Robert A. Cooper to head a group of five full-time employees to focus on company growth. The group was

provided with assistance required from idea conceptualization to its commercialization. For instance, it includes a four-day "business builder" session that helped group members generate and prioritize their different business ideas and concepts. This was followed by spending four to eight weeks period to develop the business plan of the ideas / concepts so developed. The result was the proven Market Driven Growth program at E.I. du Pont de Nemours and Co.

The Producer Model: As with the enabler and advocate models, the producer model of corporate entrepreneurship also aims at encouraging latent entrepreneurs within the organization. Accordingly, organizational members are assigned responsibility, with enough provisions for funds, to develop new ideas and concepts for business growth. To quote, Motorola has a dedicated staff of more than 35 people as corporate entrepreneurship group with an annual budget corpus in the tens of millions of dollars. The producer model, at the same time, also tries to protect emerging projects from turf battles, encourage collaboration across the units, and create pathways for entrepreneurial executives to pursue careers outside their business units. This gives emergence to independent entrepreneurs which may become threat to the organizations they left. However, the producer model is subject to suffer from certain challenges like:

First, it involves significant investments to be made over the period.

Second, integrating successful new projects developed through corporate entrepreneurship into the established business units is often found difficult.

Third, the corporate entrepreneurship group, particularly when they have pilfered top talent of organization, tend to emerge as threats to existing business units.

Now, the four models can be summarized as follows:

The Opportunist Model	Diffused ownership and *ad hoc* resource allocation
The Enabler Model	Diffused ownership and dedicated resources
The Advocate Model	Focused ownership and *ad hoc* resource allocation
The Producer Method	Focused ownership and dedicated resource

How to Nurture Corporate Entrepreneurship in India?

For corporate entrepreneurship to be successful, organizations need to nurture innovating thinking, also called, entrepreneurial mind-set among their employees. To nurture innovative thinking in organizations, management needs to provide the freedom supported by encouragement to employees capable to develop new ideas (Dess, Lumpkin and McGee 1999: 85-102). For this to happen, organizations need to take certain steps to nurture the innovative thinking among their employees. These include but are not limited to the following only:

Develop Innovative Philosophy: Innovation or creativity (i.e. thinking out of the box) needs to jell with the philosophy of the organization. So, to say, organization needs to foster creative or innovative philosophy to more and more innovations take place.

Design Flexible Organizational Structure: Innovative thinking or thinking requires facilitations like flexible organizational structure. Rigid organization structure limits rather hamper creative thinking which is the hallmark of innovation. Employees also need to be encouraged to circumvent the rigid organizational structure and procedure.

Recognize and Reward Innovative Employees: Innovative employees need to be recognized and rewarded for their innovative ideas and outcomes. This helps in two ways. *First,* the innovative employees get encouraged for their innovation and other employees get motivated to be innovative for getting recognition and reward. *Second,* this fosters spirit of innovative thinking among the employees which, in turn, improves the quality of innovation in the organization. Reward system devised should other employees to risk and achieve.

Accept Failure as Learning Experience: That failure in new experiments would not be liked rather punished by the organization stops employees from experimenting new ideas in their works. Recognizing that many good innovations have taken place after initially becoming failure, organizations should accept failures, if any, as learning experience to encourage employees to think

in innovative manner. In other words, employees should feel free to experiment new ideas without fear of failure and punishment.

Allow Bootlegging of Ideas: One way to nurture innovative thinking in the organization is allowing employees to secretly working in innovative manner all the times: be at work or off the work. In other words, employees should be allowed and encouraged to live their innovative thinking day in and day out.

Types of Corporate Innovation

Corporate innovations are classified into two broad types:

Radical Innovation: Radical innovation is also known as 'disruptive' or 'discontinuous' innovation. Radical innovation means to find out an altogether new method of doing business. Tushman and Romanelli (1985: 171-222) have defined radical innovation as "*processes of reorientation wherein patterns of consistency are fundamentally reordered.*" Radical innovation is found as a complex, difficult and even risky process. It is usually the smaller firms or start-ups that can inaugurate and introduce radical innovations.

Incremental Innovation: Incremental innovations is a change that implies small adaptions to the *status quo* (Tushman and Romanelli, 1985: 171-222), and it is often described as a step-by-step process. As a matter of fact, incremental innovation is the dominant form of innovation. A good example of incremental innovation is Google's Gmail, the world's most popular email service. It has been noticed that many a times incremental innovation takes place after a radical innovation introduces a breakthrough. It is also opined that an organization can do 1,000 things 1 percent better rather than waiting to do one thing 1,000 percent better. Research also reveals that incremental innovation is mainly found in service sector of an economy (Sundbo and Gallouj 2000:15-36).

Barriers to Corporate Entrepreneurship in India

"There is nothing more difficult to take in hand, more perilous to conduct, than to take a lead in the introduction of a new order of things, because the innovation has for enemies all those who have done well under the old conditions and lukewarm defenders in those who may do well under the new." - *Machiavelli,* The Prince

Research evidence suggest that creating new businesses in the existing corporations, or call it corporate entrepreneurship, helps increase profit margins (Wolcott Lippitz 2007). But corporate history is also witnessing that corporate entrepreneurship is littered with failures. Nonetheless, there is no denying the fact that corporations in India have realized the need for promoting entrepreneurial managers within organizations for organizational survival and growth. Nonetheless, there has not been much headway in this direction because of certain barriers to developing corporate entrepreneurship in the country. The major barriers to corporate entrepreneurship in India include:

- Resistance to Change:
- The inherent nature of large organizations
- Lack of entrepreneurial talent
- Inappropriate compensation method
- Conditions Necessary to Make Corporate Entrepreneurship Successful

Corporate entrepreneurship, like other economic initiatives, does not happen in vacuum and naturally. For corporate entrepreneurship to take place and develop needs conducive environment in the organization. The conditions necessary to make organizational environment conducive for corporate entrepreneurship to take place in India include but are not limited to the following only:

- Give people ownership
- Make risk taking and failure acceptable
- Create learning culture
- Support people with ideas

- Create a safe place for innovation
- Celebrate and reward entrepreneurial behaviour
- Encourage cross-discipline projects
- Encourage knowledge sharing
- Create and allocate funds
- Concluding Remarks

In today's highly competitive business environment, innovation is no longer a choice but compulsion for organizations to survive and thrive in the market. Innovation is a joint venture between organization and its innovative employees. Therefore, innovation to be effective, both must play their roles.

One, organizations must create innovative environment at workplace by way of extending encouragement and needed support to entrepreneurial employees to concentrate on experimenting innovations in the organizations. Experience shows nothing works like encouragement in making people to exert their best mite in performing their jobs. Here, the views of Robert D. Hishrich's et.al. (2017) seem worth citing: *"If your employees aren't innovating, your company is losing its competitive edge. Use Corporate Entrepreneurship to give your mavericks what they need to keep your company on top—all over the world."*

Two, organizational initiatives to foster corporate entrepreneurship will be effective only when employees are genuinely interested and committed to act entrepreneurially. Entrepreneurial history of India is a testimony to the fact that mere extending facilities in terms of incentives, concessions, and tax benefits failed to foster entrepreneurship in the country during sixties and seventies. The reason was that people were not ready and willing to assume entrepreneurial (i.e. risk-taking) ventures. This gave genesis to thoughtfully designed interventions called 'Entrepreneurship Development Programmes (EDP's) to prepare people to assume entrepreneurial career. EDP has by now become a movement in India and is followed by several Asian Countries in their efforts to develop entrepreneurship. Without these, any deliberation of corporate entrepreneurship will be just rhetoric and fledging a dead horse. Sooner these happen, better will be corporate entrepreneurship in India.

References

Dess, Gregory, G. T. Lumpkin and Jeffrey E. McGee (1999): Linking Corporate Entrepreneurship to Strategy, Structure, and Process: Suggested Research Directions, *Entrepreneurship Theory and Practice,* 23 (3).

Hisrich, Robert D., Michael P. Peters and Dean A. Shepherd (2017): *Entrepreneurship*, Tenth Edition, McGraw-Hill Education, New York.

Khanka, S. S. (2009): Motivational Orientation of Assamese Entrepreneurs in SSE Sector, *The Journal of Entrepreneurship,* 18 (2), June-Dec.

Pinchot, Gifford (1985): *Intrapreneurship,* Harper & Row, New York.

Sharma, Pramodita and James J. Chrisman (1999); Toward a Reconciliation of the Definitional Issues in the Field of Corporate Entrepreneurship, *Entrepreneurial Theory and Practice*, Spring.

Sundbo, J. and F. Gallouj (2000): Innovation as a Loosely Coupled System in Services. *International Journal of Services Technology and Management,* 1(1). (Eds.): *Research in Organizational Behavior, Volume* 7.

Tushman, M. L. and E. Romanelli (1985): Organizational Evolution: A Metamorphosis Model of Convergence and Reorientation, In: L. L. Cummings and M. B. Staw

Wolcott, Robert C. and Michael J. Lippitz (2007): The Four Models of Corporate Entrepreneurship, *MIT Sloan Management Review,* Vol. 49, Number 1.

arXiv:1912.12122v1 [cs.CR] 10 Dec 2019

DEEP LEARNING BASED ANDROID MALWARE DETECTION FRAMEWORK

A PREPRINT

Soumya Sourav
Department of Computer Science
University of Texas at Dallas
Dallas, TX, USA 75080
sxs180011@utdallas.edu

Devashish Khulbe
Center for Urban Science + Progress
New York University
New York, USA 10003
dk3596@nyu.edu

Naman Kapoor
Department of Computer Science
Delhi Technological University
New Delhi, India 110042
naman.kapoor12@gmail.com

December 30, 2019

ABSTRACT

With the development in the field of smartphones and ever growing base of Internet, various softwares are left prone to many malicious activities like pharming, phishing, ransomware, spam, spoofing, spyware, eavesdropping, etc. These threats have not spared the smartphones which are equally prone to them. In this work, we aim to detect these malwares with accuracy and efficiency. This being essentially a classification problem, we use various machine learning methods for this task. We observe that across models, Attention based Artificial Neural Networks (ANN), or broadly speaking, Deep Learning, are most suitable for this problem. Attention based ANNs are an amalgamation of accuracy and efficiency, the crux of our work. The accuracy achieved by our model is around 96.75%. Our model runs the test on Android Package Files (APKs) to determine whether a particular application is malicious or not by doing behavior analysis on android application under consideration.

***K*eywords** Malware detection · Deep Learning · Permissions, APK

1 Introduction

[1] The recent ubiquitousness of mobile phones for doing each and every task in day to day life because of there increased functionality has left them vulnerable to many new malware threats. Along with the nature of mobile devices, which are upgrading each day, the mobile threats are also upgrading with each passing day. The term mobile now includes all kinds of devices of IoT. Threats can be in the form of banking trojans, ransomwares, spywares, etc. which often leads to stealing private data and money from the user. Malware detection has continued to be one of the biggest challenges in today's technological world considering the fact that the resources available for doing the same are very limited. [17] According to McAfee malware analysis report 2017 the unprecedented rise of ransomwares is expected to continue in 2017. Google has removed more than 4000 apps from playstore in the past year without any information to the users. According to the Telemetry data obtained by McAfee Mobile Threat Research, more than 500,000 mobile users still have these apps installed in their devices thus making them vulnerable to the threats these apps pose . [15] Kaspersky Mobile security products detected 1,319,418 malicious installation packages, 28,796 mobile banking trojans, 200,054 mobile ransomwares. TrojanBanker has become one of the most easiest and most used malware by fraudsters.AndroidOS.Asacub is one of the malwares most prominently used. The Q2 2017 results of the Kaspersky Lab were equally threatening with malware detection being at 1,319,148 which is twice the previous quarters data. In Q2 2017, Adware with a contribution of 13.31%, was the biggest source of malwares. The share increased by 5.99%. The majority of all discovered Installation packages are detected as AdWare.AndroidOS.Ewind.iz and AdWare. AndroidOS. Agent.n. Trojan-SMS malware (6.83%) ranked second in terms of the growth rate: its contribution increased by 2.15 percentage points. The problem of using machine learning based classifier to detect malwares using android permissions presents 3 main challenges: first, we need to extract an apps main features which will provide the base to classify them; second, important features should be identified and taken into account from the obtained dataset.; third, classification using the model. The first problem can be solved using various reverse engineering tools like androguard, apk tool, etc.

which extracts an APK file to give its permissions, API calls and other related information about the file. We have used androguard in this project. The second problem is solved using Principal Component Analyses for Logistic Regression, Gradient Boosting and Naive Bayes and [4] sklearn.ensemble.ExtraTreesClassifier (ETC) for Neural Network. ETC is a meta estimator that fits a number of randomized decision trees (a.k.a. extra-trees) on various sub-samples of the dataset and use averaging to improve the predictive accuracy and control over-fitting. The third problem is solved using separate models out of which the Artificial Neural Network model that consists of an eleven-layered structure with 105 number of nodes in each layer gives the highest accuracy. This paper is organization augments understanding at every level: in Section II, we have discussed the related works in this field; in Section III, we have explained the working of our model; in Section IV, we have presented the components of our methodology; in Section V, we have discussed our experimental results; & finally, in Section VI, we have provided the references of our project .

2 Related Work

The machine learning methods for classification [1,2,3,4,5,8] has been very prevalent among data scientists and researchers. With some changes in their framework some encouraging results have been obtained. Another work [2] used the ensemble features to categorize among the malware apps. They have utilized the androguard as well for breaking down the android permissions along with mining the API calls and software/hardware features. The project is implemented in two phases: first one includes using separate features and the other one using the combined or the ensemble features. The accuracy was around 93% for the ensemble technique. One other work [4,9] has used android permissions for their projects as well. However, k-nearest classifier [4] is used in one and random forest tree classifier [9] is used in another. Both are very robust and powerful machine learning algorithms. By analyzing the number of times each system call has been issued during execution of certain task, Crowdroid [6], a machine learning-based framework, recognizes Trojan-like malware. The trojanized version of an application is very different from its genuine application, since every time it is used, it issues different types of and different number of API calls. The project [7] makes use of parallel functioning of machine algorithms for storing different features. The algorithms used are Rule Based Classifier, Function Based Classifier, Tree Based Classifier, Probabilistic Classifier. All the features from these algorithms, after they are trained, are combined and a net result is obtained whether an app is malicious or not. The proposed method generates some good results, however the inclusion of four different algorithms surely speeds it down.

3 Methodology

3.1 Part 1: Data & Features

First phase of the work involves reading dataset, which includes permissions and API calls of various apps which are classified whether they are malwares or benign applications. Relevant features are then selected using two separate techniques: Principal Component Analyses (PCA) and a wrapper classifier by segregating features w.r.t. their importance. Irrelevant features are then removed for better generalization and better accuracy as well as reduced time involved in classification.

3.1.1 Dataset

The dataset used for analysis is provided by Kaggle, a competitive platform where data miners and scientists can obtain numerous datasets and can compete in various challenges hosted by this domain only. Generally companies and researchers post some data on it and data miners compete to provide the best predicting accuracy on that data. The dataset contains features of 338 applications which are labelled as ‘benign’ or ‘malicious’. This label is used to do feature selection and classification analysis by supervised learning. The dataset comprises of a .csv file. The top row contains all the permissions. The presence of a specific permission in an application sample is indicated by 1 and 0 if it is not present.

Imbalance problem - the situation when the contents of each class are not proportionate- is a very important aspect when using binary classifiers for the detection of malicious code. In our case the absence of malwares and legitimate files in equal proportions causes the imbalance problem. It is need of hour to continuously update the training set with new malicious files for making the classifier better. This is really an important aspect in maintaining such a framework. The dataset is updated to accommodate for new samples using androguard, an opensource project to extract features from APKs. This solves for the imbalance problem by maintaining the proportion of benign and malicious entries to be equal.

Type	Features
API call related	abortBroadcast, getDeviceId, getSubscriberId, getCallState, getSimSerialNumber, android.provider.Contacts, android.provider.ContactsContract; HttpUriRequest, SMSReceiver, bindService, onActivityResult, LjavaxCryptoSpecSecret, DexClassLoader, getNetworkOperator, getSimOperator
Command related	.apk; pmsetInstallLocation; pminstall; GET-METADATA; GET-RECEIVERS; GET-SERVICES; GET-SIGNATURES; GET-PERMISSIONS
Permissions	android.permission.UPDATE-LOCK; android.permission.USER-ACTIVITY; android.permission.VIBRATE; android.permission.WAKE-LOCK; android.permission.WRITE-CALENDAR; android.permission.WRITE-CALL-LOG; android.permission.WRITE-CONTACTS; android.permission.ACCESS-FINE-LOCATION

Table 1: Overview of the features extracted from the apps.

3.1.2 Features

A feature is a significant estimable property or characteristic of an event under observation. Each attribute in the dataset set signifies a unique feature and each entity signifies a data sample (an android application in this case). A feature in this project corresponds to a permission specified by an application. The dataset contains 330 features. The feature vector contains feature data in binary form where '1' indicates the presence of a permission and '0' indicates absence. The malware analysis approach implemented in this paper uses static android features extracted from the android app which are used to determine if app contains any malicious permissions which depends on a trained classification model. Androguard does the job of extracting features to train the ANN model combination of malware and benign APKs. Three categories of features are used for learning phase: 1. App Permissions 2. API calls 3. Standard OS and framework commands. API calls are the calls which are made to the server in the name of an application using a SDK or an API. Android permissions are requests or permissions acquired by the apps in order to use certain system data and features to maintain security for the system and user.

3.1.3 Feature Extraction

Overfitting is an issue where the model instead of describing the underlying relationship, describes noise and random error. An overfitted model tends to perform very poorly on predictive tasks as it overreacts to minor fluctuations in training data. This is a very common problem with every machine learning model. Huge and complex datasets affect training models by reducing the efficiency of the classification problem by using unnecessary number of features resulting in increased computation time. Hence extraction of important features is a key part before modeling. We use two methods for selection of important features:

Principal Component Analysis Principal Component Analysis (PCA) [12] is an important technique to deal with multicollinearity in the data and eliminating redundant variables. It drops the "least important" variables while still retaining the most valuable parts of all of the variables. As an added benefit, each of the new variables after PCA are all independent of one another. In our data set, the original feature space of 330 variables is reduced to 23 variables after using PCA. This transformed feature space accounts for 90% of the variance in the original space.

ExtraTreeClassifier ExtraTreeClassifier is an amalgamation of a search technique, which selects the features, and an evaluation measure, used to score the feature subsets. The algorithm varies in complexity with some being as simple as testing all the possible subsets of selected features and then winnowing the one with minimal error rate. One of the most influential factor for the algorithm is the choice of evaluation metric and these evaluation metrics are the one's which create a distinction among three main feature selection algorithm: filters, embedded methods and wrappers. This technique makes use of wrapper method to figure pout the accuracy of the selected features' subsets. By counting the number of mistakes our model makes while predicting on the hold-out set, the accuracy can easily be determined.

Wrapper methods tend to be very intensive when it comes to computations. However the best results on a model can be obtained by this method only.

3.2 Part 2: Modeling

Second part involves training the machine learning models using the processed dataset and saving weights and the trained model. We use Logistic Regression, Naive Bayes and Gradient Boosted Trees along with the artificial neural network and compare their performance on the test data. We use the truncated feature space after the PCA and ExtraTreeClassifier segregates the useful variables. The test dataset is dynamic and keeps on changing whenever the program is run. It consists of around 35-40 permissions belonging to the original dataset with almost 330 permissions and generated on 398 applications both benign and malicious. This models after being trained, saves the weights and the model. These weights and model are then called which breaks individual apps to be checked and then the neural network model decides whether the app is malicious or benign.
The crux of our work is developing the Artificial Neural Network for this problem and compare it with other off the shelf techniques. Thus, we focus on the results of ANN for major parts of the modeling and results section.

3.2.1 Classification Methods

We begin with off the shelf models for the classification task. We use Logistic Regression, Naive Bayes, Random Forests and Gradient Boosting models before moving onto deep neural networks. These methods are fairly well known and have proven to perform good for binary classification problems. The problem with these approaches, however, is that they might not be able to learn complex non-linearity which might be in the data. To tackle this, we delve into developing Artificial Neural Network for the task.

For the Artificial Neural Network we add an attention layer to enhance the prediction capability of the neural network. The neural architecture has 12 Dense layers. In between every third Dense layer a dropout layer was added to reduce the overfitting of the neural networks. The dropout layer drops a certain percentage of the neurons in random. These neurons don't receive any updates and don't contribute to the output at all. The attention network [13, 14] is used in order to make the neural network capable of attending to a subset of the inputs which increases the space of functions which can be approximated by the neural network and makes it possible to look into entirely new use-cases which enhances the neural net's capability to converge to a global minimum. The attention network generally assigns the scores to the input features using the softmax function.

$$f(z)_{\mathrm{j}} = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_i}} \tag{1}$$

where f_j is the output of attention layer and e_j is the output previous Dense layer. This output is the probability distribution which is combined with the Dense layer output which gives the importance of each feature for the output prediction. Different methods for combining the output layer and probability distribution can be used like Dot product, Scalar Multiplication or Bilinear combination. For the final training procedure of the PKG and PCB model, a batch size of 32 was fixed after doing cross validation with the optimizer "Adam" was used.

4 Results

For analyzing the performance of our models, we decided to check their accuracies on the test data. Checking the accuracy makes sense as the distribution of positive and negative records in our data are equal in proportion. The training:test data proportion is set to be 75:25.

On comparing the accuracy across all models, ANN, due to its complex structure, is able to perform better than all of the other classification techniques. The validation accuracy is found to be 95%. The problem of classification could be countered with many machine learning algorithms like SVM or regression. However, the main advantage Neural Networks or the multilayer perceptrons hold over these algorithms are the fixed number of layer size. The neural networks are parametric models while the machine learning algorithms are not i.e. in neural networks, there are a number of hidden layers depending upon the number of features. The outputs of neural networks can be multiple and they were tested to be faster than the machine learning algorithms.

We also check the performance of our ANN model with respect to its predictions concerning the individual classes through a confusion matrix.

Precision is a common metric to judge the performance of classification models. It is essentially the fraction of relevant examples classified correctly. With the ANN model, the precision score is found to be 0.90476.

Name	Accuracy(%)
Naive Bayes	88.3
Random Forests	90.8
Gradient Boosting	91
Logistic Regression	93.6
Attention Based ANN	95.3

Table 2: Classification accuracy across models

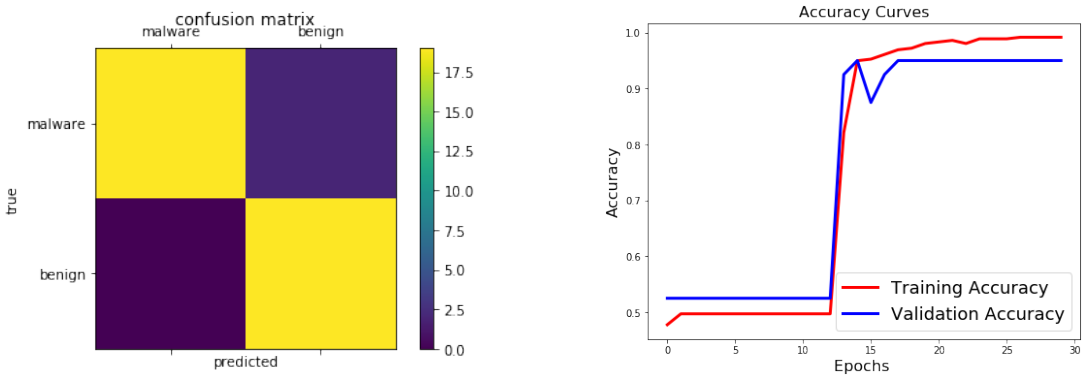


Figure 1: Performance of ANN model: Confusion matrix (left) and accuracy curve (right)

$$ANNPrecision = \frac{TruePositive}{TruePositive + FalsePositive} = 0.90476 \tag{2}$$

5 Conclusion

We introduced a new Deep Learning based method to predict malwares in Android applications which performs significantly better than traditional off the shelf machine learning techniques. Due to the small size of the data we worked on, this new method is yet to be tested on bigger and more complex data with more features and across various platforms. This can be a possible scope for future work. Also, we fed only the significant features from our original data into the models' inputs. There can be further scope in finding which features are important predictors for this task. Furthermore, we think malware detection is a fairly important area of study and different approaches and data sets still need to be explored in this domain.

References

[1] Justin Sahs, Latifur Khan. "A Machine Learning Approach to Android Malware Detection", European Intelligence and Security Informatics Conference-2012

[2] A.M. Ashwini, P. Vinod. "Android malware analysis using ensemble features", Springer international publishing Switzerland 2104

[3] Nikola Milosevic, Ali Dehghantanha, Kim-Kwang Raymond Choo. "Machine learning aided Android malware classification", Elsevier ltd, 2017

[4] Naser Peiravian, Xing Quan Zhu. "Machine learning for android malware detection using permission and api calls", IEEE 25th international conference on tools with artificial intelligence 2013

[5] I. Burguera, U.Z., Nadijm-Tehrani. "Crowdroid: Behavior- Based Malware Detection System for Android", SPSM'11, ACM(October 2011)

[6] Suleiman Y. Yerima, Sakir Sezer, Igor Muttik. "Android Malware Detection Using Parallel Machine Learning Classifiers",8th International Conference on Next Generation Mobile Applications, Services and Technologies, (NGMAST 2014), 10-14 Sept., 2014

[7] Zami Aung, Win Zaw. "Permission-based android malware detection", international journal of scientific and technology research, vol. 2, issue 3 , 2013

[8] Benjamin VA, Chen. "Machine learning for attack vector identification in malicious source code", 2013 IEEE international conference on intelligence and security informatics (ISI). IEEE; 2013. p. 21–3

[9] Boxall A. The number of smartphone users in the world is expected to reach a giant 6.1 billion by 2020; 2015 http://www.digitaltrends.com/mobile/ smartphone- users- number- 6- 1- billion- by-2020/

[10] Schmidt A-D, Clausen JH ,Camtepe A ,Albayrak S. "Detecting symbian os malware through static function call analysis", 4th international conference on malicious and unwanted software (MALWARE), 2009. IEEE; 2009. p. 15–22

[11] Alam MS, Vuong ST. "Random forest classification for detecting android malware", 2013 IEEE and internet of things (iThings/CPSCom), IEEE international conference on and IEEE cyber, physical and social computing, green computing and communications (GreenCom). IEEE; 2013. p. 663–9

[12] Wold, S., Esbensen, K. and Geladi, P., 1987. Principal component analysis. Chemometrics and intelligent laboratory systems, 2(1-3), pp.37-52.

[13] https://skymind.ai/wiki/attention-mechanism-memory-network

[14] http://akosiorek.github.io/ml/2017/10/14/visual-attention.html George Kour and Raid Saabne. Real-time segmentation of on-line handwritten arabic script. In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 417–422. IEEE, 2014.

[15] Guy Hadash, Einat Kermany, Boaz Carmeli, Ofer Lavi, George Kour, and Alon Jacovi. Estimate and replace: A novel approach to integrating deep neural networks with existing applications. *arXiv preprint arXiv:1804.09028*, 2018.

Wireless Personal Communications
https://doi.org/10.1007/s11277-020-07078-8

Design and Analysis of a Multi-band Flower Shaped Patch Antenna for WLAN/WiMAX/ISM Band Applications

Karteek Viswanadha[1] · **Nallanthighal Srinivasa Raghava**[1]

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract
This paper presents the design and parametric analysis of a flower-shaped patch antenna. The proposed patch antenna is designed using FR-4 substrate mounted on the Jerusalem cross-shaped DGS structure. A Double Negative (DNG) triple Complementary Split Ring Resonator is embedded inside the substrate. A circular foam substrate with the dimensions of $10\times4\times10\times2$ mm^3 is sandwiched between the patch and the FR-4 substrate. The overall dimensions of the patch are 23.5×16 mm^2. The proposed antenna resonates at 5.2 GHz and 8.25 GHz respectively. This also possesses wide bandwidth of 1.2 GHz (24.1%) in the range of 4.95–6.15 GHz and 2.2 GHz (26.5%) in the range of 7.1–9.3 GHz. The gains in these bands are observed to be 3.93 dBi and 5.02 dBi respectively. The whole design is carried out in CST-microwave studio 2016 utilizing Finite Integration Technique. The developed multiband antenna can be useful for several wireless communication applications, such as WLAN, WiMAX, and ISM band. The proposed antenna is fabricated and its performance parameters are measured. The simulated and measured results are in good agreement.

Keywords Flower patch · Jerusalem cross DGS · EBG · Surface waves · WLAN · WiMAX

1 Introduction

Microstrip patch antennas can easily be modified depending on the system specifications and the type of application. Different shapes of patch antennas are proposed along with the Electromagnetic Bandgap structures (EBG) in order to miniaturize the antenna along with the improvement in the performance parameters of the later. In [1], stacked patch antennas along with the rectangular lattice holes on the ground plane is proposed. The proposed patch antenna in [1] resonates at 2.45 GHz with the improved efficiency of 96.53%. A slotted stacked E-shaped stacked patch antenna is proposed in [2]. The antenna in [2] possesses gain of 12 dBi at 2.84 GHz with maximum efficiency

✉ Karteek Viswanadha
karteekviswanath@gmail.com

Nallanthighal Srinivasa Raghava
nsraghava@dce.ac.in

[1] Delhi Technological University, New Delhi, India

Published online: 21 January 2020

of 99.98%. A stacked square shaped patch antenna with shorting post with a reflector at 8.5 mm from the ground plane is proposed in [3]. A squared slotted EBG is proposed in [3]. The proposed antenna in [3] resonates at 2.63 GHz with the gain and directivity of 4.63 dBi and 4.78 dBi. Apart from the rectangular shape patch antenna, different shapes of patches are proposed in order to achieve high gain and bandwidth along with the compactness. One such type of patch is circular patch. This patch is compact and provides high gain. A circular patch antenna is designed to work at 2.45 GHz with the reflection of − 10.199 dB and VSWR of 1.6 at the resonant frequency in [4]. 8-element circular patch array is designed to work at 3 GHz with the gain of 11 dB and bandwidth of 166 MHz in [5]. A circular slotted patch antenna with the dimensions of $28 \times 16 \times 0.8$ mm^3 is proposed to work at 2.5/3.5/5.7 GHz in [6]. The antenna in [6] possesses a very narrow bandwidth in all the bands. The circular patch antenna with inverted 'T' slot working at 2.4 GHz and 5.5 GHz having a radius of 11.4 mm and dimensions of 30 mm × 30 mm × 1 mm^3 is proposed in [7]. The dimensions of the slot are 2 mm × 16 mm and 2 mm × 13.5 mm. The directivities and bandwidths of the antenna in [7] are 4.78 dB, 7.26 dB, 80 MHz, and 510 MHz respectively. In [8], the printed parts of the versatile antenna are used in different operating modes by PIN diodes. This feature makes the antenna work over a wide range of frequencies. The reflector array of 421 circular elements with semicircular ring slots in [9]. The proposed array in [9] possesses a gain of 25.2 dB at 15 GHz. A distorted form of circular patch known as elliptical patch antenna is also widely used in order to achieve wide bandwidth. The elliptical patch with an elliptical notch in the ground is proposed in [10]. The proposed antenna resonates at 1.71 GHz, 2.28 GHz, 3.03 GHz and 3.84 GHz with the bandwidth of 2.78 GHz. The gain of the antenna in [10] varies from 2.78 to 5.52 dB. An antenna with different slots is proposed in [11]. The antenna in [11] possesses a gain and bandwidth of 6.5 dB and 1.31 GHz in the band of 8.9 – 10.21 GHz. A butterfly-shaped antenna with rectangular slots is proposed in [12]. The antenna achieved a gain of 7.22 dB in the band of 6.02–7.47 GHz with the centre frequency of 6.45 GHz. The overall dimensions of the antenna are $40 \times 40 \times 3$ mm^3. The antenna works in the band of 3.7–5.45 GHz in [12]. It achieves axial ratio bandwidth of 38% in the above said band. In [13], a novel circular ring patch antenna is proposed. The antenna works at 6.68 GHz, 12.68 GHz, 13.75 GHz and, 14.2 GHz respectively. The bandwidths in these ranges are obtained as 270 MHz and 1 GHz respectively. A novel fractal antenna with the dimensions of 60 mm × 60 mm × 1.6 mm^3 is proposed in [14]. The antenna is loaded with circular shaped Complementary Split Ring Resonator (CSRR). A peak gain of 7.7 dB is obtained with very narrow bandwidth in [15].

A dollar shaped patch is proposed in [15]. The antenna resonates at 2.2 GHz and 8.8 GHz respectively. The bandwidths of the proposed antenna in [15] are observed to be 1.1 GHz and 2.4 GHz respectively. A hexadecagon shaped circular patch antenna is proposed in [16]. The antenna resonates at 13.67 GHz and 15.28 GHz respectively. The impedance bandwidths are observed to be 0.854 GHz and 1.1 GHz respectively. The proposed choice of metamaterials often miniaturizes the antenna which is indeed a boon to the field of wireless communications. Different shapes like 'H' are proposed and analysed in order to optimise the antenna [17]. In [18], a 4-segment SRR metamaterial with Mu Negative is proposed which enhances bandwidth by 30 GHz, directivity by 2% and front to back lobe ratio by 9%. A 'C'-shaped pair metamaterial is proposed to create resonance at C-band with the increment in gain and directivity along with the reduction in antenna size in [19]. In [20], a pair of capacitively loaded metamaterial with Double Negative (DNG) is proposed to resonate at 8.5, 17.7, 20, 23.7 and 30 GHz with the 72% miniaturization of the antenna. Different shapes of antennas with

modified Defective ground, Electromagnetic Bandgap Structures (EBG) and Metamaterials (MTMs) are proposed in [21–32] with wideband bandwidth, small frequency ratio, high gain and directivity.

This paper presents the design and parametric analysis of the flower shaped patch antenna loaded with triple CSRR and Jerusalem cross-shaped DGS at 5.2 GHz and 8.25 GHz. The usage of both metamaterial and DGS enhances the bandwidth and gain of the proposed antenna. The proposed patch has efficiencies of 61% and 69.7% at 5.2 GHz and 8.25 GHz respectively.

The paper consists of four sections. In Sect. 2 the evolution of the proposed antenna structure along with corresponding return loss plots are presented. The mathematical analysis is in Sect. 3. The detailed parametric analysis is presented in Sect. 4. All the simulated, measured results and comparison table are presented in Sect. 5.

2 Evolution and the Design of the Proposed Antenna

The proposed patch is designed using copper with the thickness of 35 µm. The whole antenna is mounted on FR-4 (lossy) substrate (ε_r=4.4) with the thickness of 1.5 mm. 8.25 GHz and 5.2 GHz are the two frequencies of interest. For $f_r^{'}$=8.15 GHz, the corresponding effective radius and the actual radius is given by

$$f_r^{'} = \frac{1.8412 * c}{2\pi(r_e)\sqrt{\varepsilon_r}} \tag{1}$$

where 'c' is the velocity of light and 'ε_r' is the relative dielectric constant.

$$r_e = r\left[1 + \frac{2h}{\pi r \varepsilon_r}\left\{\ln\left(\frac{r}{2h}\right) + (1.41\varepsilon_r + 1.77) + \frac{h}{-}r(0.268\varepsilon_r + 1.65)\right\}\right]^{\frac{1}{2}} \tag{2}$$

From the above two equations, effective radius is obtained as 8.25 mm and the corresponding radius (r) is obtained as 7.5 mm. Similarly, for the frequency $f_r^{''}$=5.15 GHz, the effective and the actual radii are given by:

$$f_r^{''} = \frac{1.8412 * c}{2\pi(r_{2e} + r_{1e})\sqrt{\varepsilon_r}} \tag{3}$$

$$r_{1e} + r_{2e} = (r_1 + r_2)\left[1 + \frac{2h}{\pi(r_1 + r_2)\varepsilon_r}\left\{\ln\left(\frac{(r_1 + r_2)}{2h}\right) + (1.41\varepsilon_r + 1.77) + \frac{h}{(r_1 + r_2)}(0.268\varepsilon_r + 1.65)\right\}\right]^{\frac{1}{2}} \tag{4}$$

where r,r_1 and r_2 is the radius of the central and surrounding patches and 'h' is the height of the substrate. The proposed patch is derived from basic geometric theorem stating that if three circles of radii r, r_1 and r_2 (where $r > r_1 > r_2$) touch externally then $\frac{1}{\sqrt{r_2}} = \frac{1}{\sqrt{r}} + \frac{1}{\sqrt{r_2}}$. By fixing r_1=2.5 mm and substituting r=7.5 mm obtained from Eq. (2), the radius (r_2) of the small patch is obtained as 1.5 mm. Figure 1 shows the designed flower shaped patch antenna. The whole design is carried out in Computer Simulation Technology microwave studio-2016 (CST-2016). The proposed antenna is dual fed with the feed lengths (l, l_1, l_2 and l_3) 4 mm, 5 mm, 11 mm and 6 mm respectively. The widths (w, w_1, w_2 and w_3) of the feed lines are 2 mm, 2 mm, 3 mm and 3 mm respectively. The whole patch is mounted on

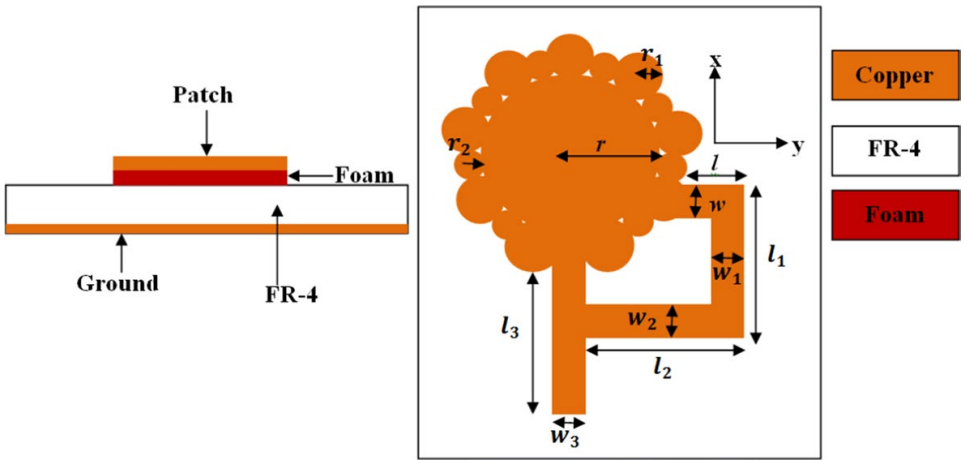


Fig. 1 Proposed antenna

the ground with dimensions of 30 mm × 30 mm and having Jerusalem cross-shaped slot. A foam substrate ($\varepsilon_r = 1.07$) with the dimensions of $10 \times 4 \times 10 \times 2$ mm^3 is introduced between the patch and the FR-4 substrate. The effects of the width of the foam on the reflection coefficient of the proposed patch are discussed in subsequent sections.

By adjusting the dimensions of the feed lines, the required frequencies are obtained and optimized independently. Table 1 shows the detailed dimensions of the proposed antenna

Figure 2a shows the different stages in the evolution of the proposed antennas and the corresponding simulated return losses are shown in Fig. 2b. The antenna presented in stage 1 of Fig. 2a is a basic circular patch antenna. From Fig. 2b (thick blue line), the resonance frequency of the antenna presented in stage 1 is about 9.62 GHz and the impedance bandwidth ($S_{11} < -10$ dB) is about 110 MHz. When only small circular patches are added on the circumference of the circular patch of stage 1, three resonant frequencies namely 4.1 GHz, 7.34 GHz and 8.1 GHz are observed with the impedance bandwidths 242 MHz, 125 MHz and 620 MHz respectively, as shown in Fig. 2b (Red line). When only big circular patches are added on the circumference of the circular patch of stage 1, three resonant frequencies namely 4.57 GHz, 6.25 GHz and 8.82 GHz are observed with the impedance bandwidths of 112 MHz, 223 MHz and 312 MHz respectively, as shown in Fig. 2b (Green line). When the small and big circular patches are added on the circumference on the circular patch of stage 1, two resonant frequencies namely 5.52 GHz and 9.15 GHz are observed with the impedance bandwidths of 100 MHz and 405 MHz respectively, as shown in Fig. 2b (Dotted purple line). Finally, dual feed lines are added to the antenna in stage 4 and two resonant frequencies namely 4.57 GHz and 5.15 GHz are observed with the impedance bandwidths of 1.25 GHz and 2.2 GHz respectively, as shown in Fig. 2b (Red dotted line). Jerusalem cross-shaped slot is treated as two perpendicular dumbbells. Dumbbells are equivalent to the parallel L-C tank circuit. A slot etched on the ground plane adds slot resistance, capacitance and inductance and thereby changing the effective inductance and capacitance. This not only reduces return loss but also harmonics. Figure 3 shows the proposed Jerusalem cross DGS. The overall dimensions (l_o and w_o) of the cross are chosen to be 15 mm × 15 mm. The lengths (l' and l'') of the arms are chosen to

Table 1 Parameters of the proposed antenna (see Fig. 1)

Parameters	r	r_1	r_2	l	l_1	l_2	l_3	w	w_1	w_2	w_3
Values (mm)	7.5	2.5	1.5	4	5	11	6	2	2	3	3

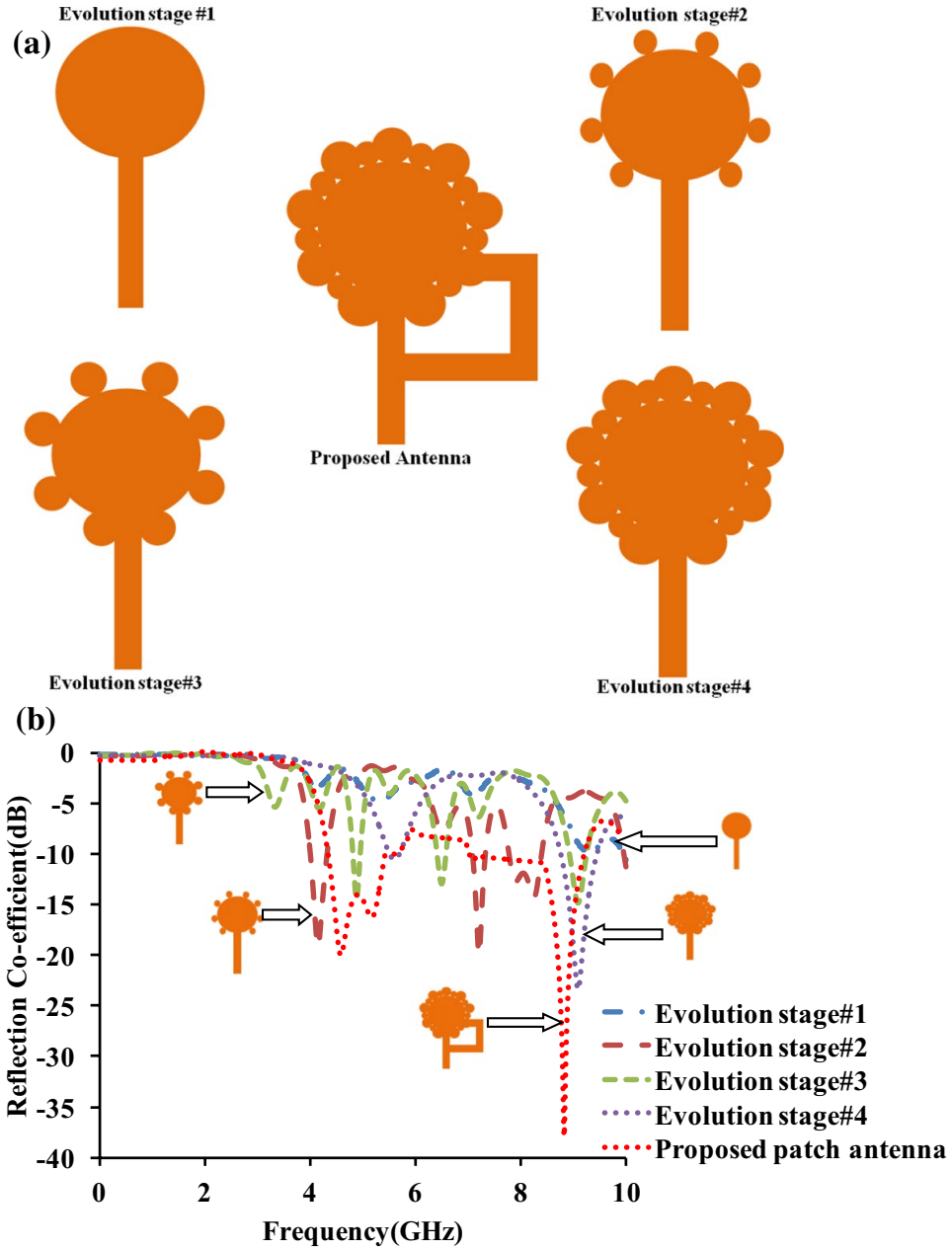


Fig. 2 **a** Evolution stages of the proposed antenna. **b** Simulated return losses of the evolution stages

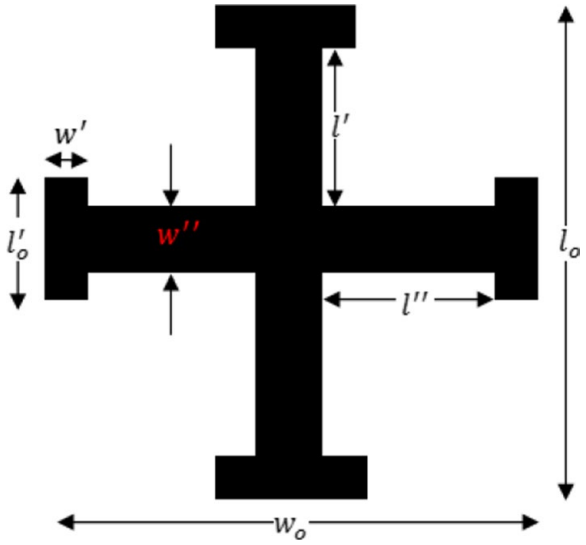


Fig. 3 Proposed Jerusalem cross shape DGS

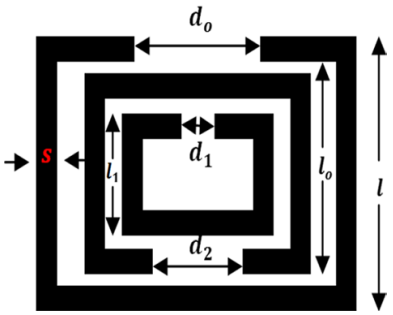


Fig. 4 An unit cell of the triple CSRR metamaterial

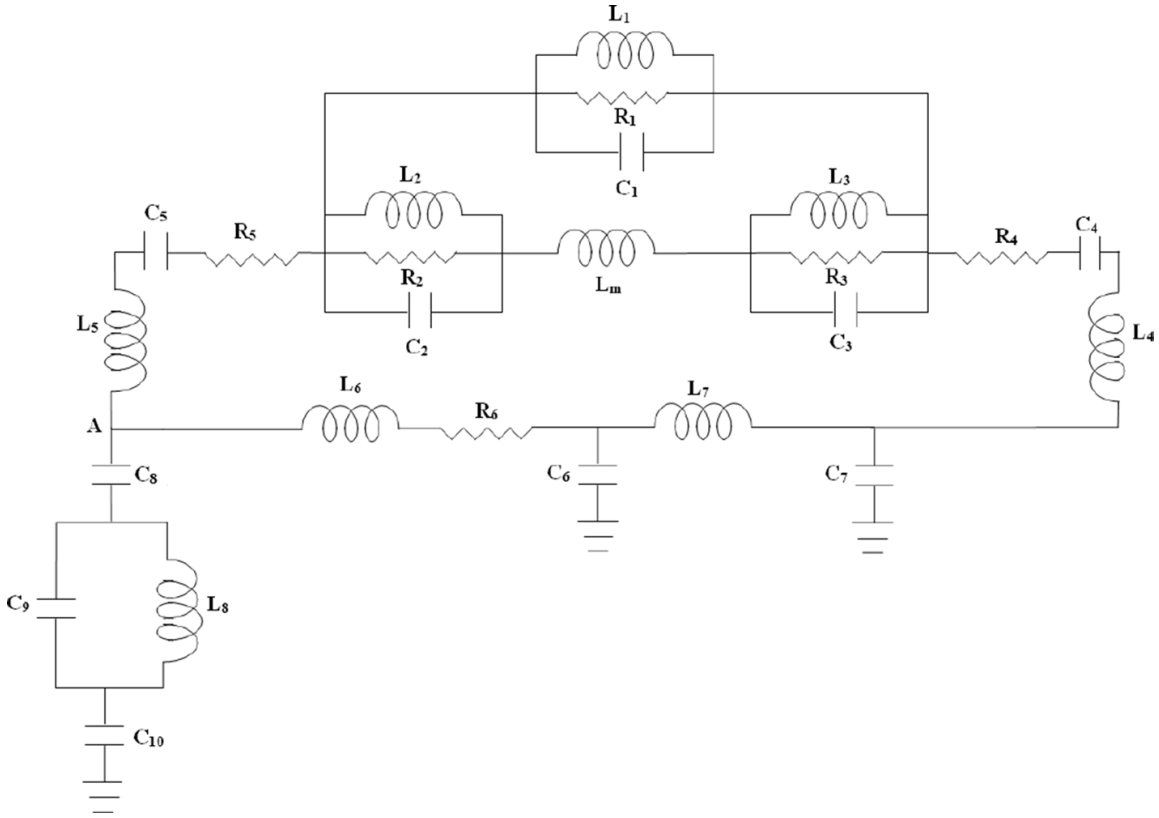


Fig. 5 Equivalent circuit of the proposed patch antenna

be 10 mm each and widths (w'') of the arms are chosen to be 5 mm each. The lengths (l'_o) and widths (w') of the end arms of the cross are chosen to be 7 mm and 2.5 mm respectively.

Figure 4 shows the unit cell of the triple CSRR metamaterials. The lengths (l, l_o, and l_1) of outer, middle and inner rings are 3 mm, 2 mm and 1.5 mm respectively. The width (s) of each ring is 0.25 mm and widths of openings (d_o, d_1 and d_2) of the rings are 1 mm, 0.75 mm and 0.5 mm respectively.

The horizontal and vertical spacing between the cells is chosen to be 5 mm each. A total of 25 cells are used as the cell dimensions and spacing are proportional to the gain and bandwidth of the patch antenna. These cells are embedded inside the substrate at the height of 0.762 mm and iterations are carried out in order to obtain good results. All the iterations are not reported in this paper as they produced unacceptable results. However, the parametric analysis on the metamaterials is carried out in the subsequent sections. Figure 5 shows the lumped element equivalent of the proposed patch antenna. The values of lumped elements are extracted from the bandwidths and resonant frequencies of the proposed patch antenna. Parallel RLC circuits on the upper branch represent the circular patches as shown

in Fig. 1. The upper RLC circuit(R_1-L_1-C_1) represents the central circular patch having the radius of 7.5 mm, the lower RLC circuits ((R_2-L_2-C_2) & (R_3-L_3-C_3)) represent the circular patches with the radii of 1.5 mm and 2.5 mm respectively. The resonant frequency of the upper RLC circuit is given by:

$$f_r' = \frac{1}{2\pi\sqrt{L_1 C_1}}$$

$$\text{Bandwidth (BW)} = \frac{R_1}{2\pi L_1}$$

The upper RLC circuit is designed to resonate at 8.25 GHz with the bandwidth of 2.2 GHz.

The resistance of R_1 is fixed at 1 Ω. The value of R_1 is chosen to be very low i,e 1 Ω in order to obtain the quality factor between 1 and 10. The quality factor in this range makes the upper RLC circuit wideband. The capacitance of C_1 and inductance of L_1 are obtained as 5 pF and 73 pH respectively. The corresponding quality factor is given by $Q_1 = \frac{1}{R_1}\sqrt{\frac{L_1}{C_1}}$

Similarly, the right most RLC circuit (R_3-L_3-C_3) is designed to work at 5.25 GHz with the bandwidth of 1.2 GHz. The bandwidth of the right most RLC circuit is given by:

$$\text{Bandwidth (BW)} = \frac{R_3}{2\pi L_3}$$

The resonant frequency of the right most RLC circuit is given by:

$$f_r'' = \frac{1}{2\pi\sqrt{L_3 C_3}}$$

The value of R_3 is chosen to be very low i.e. 1 Ω in order to obtain the quality factor between 1 and 10. The quality factor in this range makes the rightmost RLC circuit wideband. The inductance of L_3 and the capacitance of C_3 are obtained as 132pH and 7pF respectively. The corresponding quality factor is given by $Q_3 = \frac{1}{R_3}\sqrt{\frac{L_3}{C_3}}$.

Inductance is directly proportional to the square of the radius of the circular patch. Using the same concept, the value of the inductance of L_2 for the left most RLC circuit is obtained as 47.32pH. The value of the resistor (R_2) is same i.e. 1 Ω. The value of the capacitance C_2 is obtained from the formula given by:

$$f_r''' = \frac{1}{2\pi\sqrt{L_2 C_2}}$$

The value of C_2 is obtained as 19.4 pH.

The two resonant tanks ((R_2-L_2-C_2) & (R_3-L_3-C_3)) are designed to operate at 5.25 GHz. By choosing different values of L_2, C_2, L_3 & C_3 will produce wideband bandwidth in the frequency range of 4.55–5.82 GHz. The two resonant tanks are designed to work at 5.25 GHz. The same design holds for the patches with the radii of 1.5 mm and 2.5 mm where the patches are designed using the Eq. (3) to resonate at 5.25 GHz. There is a gap formed between the patches on the circumference of the center patch. The gap denotes a mutual inductance with a small coupling factor. A tight coupling between the patches leads to the shift in frequency of interest and deterioration of the return losses.

Low coupling factor decreases the bandwidth with the deterioration in the return losses. Hence, those results are not reported in this paper. A coupling factor (k) of 0.5 is chosen in order to achieve the best results.

$$L_m = k\sqrt{L_2 L_3}$$

The lumped element values of the corresponding microstrip lines (R_5-L_5-C_5 and C_8) are calculated using the equations in [21].

$$L_5 = 2l_3 \ln\left[\frac{5.98h}{0.8w_3 + t}\right] \tag{a}$$

$$C_5 = C_8 = \frac{0.264l_3\left(\varepsilon_r + 1.41\right)}{\ln\left[\frac{5.98h}{0.8w_3+t}\right]} \tag{b}$$

where l_3 = length of the trace in mm, h = height of the substrate in mm, w_3 = width of the trace in mm, t = thickness of the conductor in mm, ε_r = relative dielectric constant of the substrate.

The values are further optimized using the Agilent ADS parameter tuning feature in order achieve the wide bands in the frequency ranges of 4.95–6.15 GHz and 7.1–9.3 GHz respectively. The section (R_4-L_4-C_4-R_6-L_6-C_6-L_7-C_7) represents the right angle bend of the microstrip line as shown in the Fig. 1. The lumped element values of the right angle bends of the microstrip line are calculated using equations in [21].

$$\frac{C_4}{w} = \frac{C_6}{w} = \frac{C_7}{w} = \left(9.5\varepsilon_r + 1.41\right)\frac{w}{h} + 5.2\varepsilon_r + 7 \tag{c}$$

$$\frac{L_6}{h} = \frac{L_7}{h} = 1004\sqrt{\frac{w}{h}} - 4.21 \tag{d}$$

Table 2 shows the values of the lumped element equivalent circuit of the proposed antenna

The impedance is observed at point 'A' in the Fig. 5. The impedance transfer function $Z_f(s)$ is obtained as:

$$Z_f(s) = \frac{Z_m(s)\left[s^2\left(C_8C_{10}L_8 + L_8C_9\left(C_8 + C_{10}\right)\right) + C_8 + C_{10}\right]}{Z_m(s)C_8C_{10}s^2 + s^2\left(C_8C_{10}L_8 + L_8C_9\left(C_8 + C_{10}\right)\right) + C_8 + C_{10}}$$

where, $Z_m(s)$ represents the impedance transfer function of the flower patch antenna. The transfer function $Z_m(s)$ is obtained as below.

Table 2 Detailed values of the lumped elements

Lumped parameter	Value	Lumped parameter	Value
L_1	73 pH	C_6	13.68 pF
L_2	47.32 pH	C_7	4.98 pF
L_3	132 pH	C_8	4.48 pF
L_4	60 pH	C_9	1 pF
L_5	60 pH	C_{10}	13.475 pF
L_6	14.65 pH	R_1	0.1 Ω
L_7	190.92 pH	R_2	0.1 Ω
L_8	311 pH	R_3	0.1 Ω
L_m	39.5 pH	R_4	20 Ω
C_1	5 pF	R_5	50 Ω
C_2	19.4 pF	R_6	252 Ω
C_3	7 pF		
C_4	0.5 pF		
C_5	40 pF		

$$
Z_m(s) = \frac{\left[\begin{array}{c}
\left[\begin{array}{c} L_7L_6C_7C_6s^3 + R_6L_7C_6C_7s^2 + s\left[\begin{array}{c} L_7C_7 \\ +L_6(C_6+C_7)\end{array}\right] \\ +R_6(C_6+C_7)\end{array}\right]\left[\begin{array}{c} L_7C_7C_6s^2 \\ +(C_6+C_7)\end{array}\right]^2 (R_1L_1C_4C_5s^2)[R_2L_2(R_3L_3C_3s^2+L_3s+R_3) \\
+R_3L_3(R_2L_2C_2s^2+L_2s+R_2)+L_m(R_2L_2C_2s^2+L_2s+R_2)(R_3L_3C_3s^2+L_3s+R_3)] \\
+[(R_4+R_5)C_4C_5s+(L_4+L_5)C_4C_5s^2+(C_4+C_5)] \\
[(R_1L_1C_1s^2+L_1s+R_1)[R_2L_2(R_3L_3C_3s^2+L_3s+R_3)+R_3L_3(R_2L_2C_2s^2+L_2s+R_2) \\
+L_m(R_2L_2C_2s^2+L_2s+R_2)(R_3L_3C_3s^2+L_3s+R_3)+ \\
R_1L_1(R_2L_2C_2s^2+L_2s+R_2)(R_3L_3C_3s^2+L_3s+R_3)]+ \\
[(L_7C_4C_5C_6s^2)(R_1L_1C_1s^2+L_1s+R_1)[R_2L_2(R_3L_3C_3s^2+L_3s+R_3)+R_3L_3(R_2L_2C_2s^2+L_2s+R_2) \\
+L_m(R_2L_2C_2s^2+L_2s+R_2)(R_3L_3C_3s^2+L_3s+R_3)+R_1L_1(R_2L_2C_2s^2+L_2s+R_2)(R_3L_3C_3s^2+L_3s+R_3)]+ \\
[(L_7C_6C_7s^2)+(C_6+C_7)]^2(R_1L_1C_4C_5s^2) \\
[R_2L_2(R_3L_3C_3s^2+L_3s+R_3)+R_3L_3(R_2L_2C_2s^2+L_2s+R_2)+L_m(R_2L_2C_2s^2+L_2s+R_2)(R_3L_3C_3s^2+L_3s+R_3)] \\
+[(R_4+R_5)C_4C_5s+(L_4+L_5)C_4C_5s^2+(C_4+C_5)][(R_1L_1C_1s^2+L_1s+R_1)[R_2L_2(R_3L_3C_3s^2+L_3s+R_3)+ \\
R_3L_3(R_2L_2C_2s^2+L_2s+R_2)+L_m(R_2L_2C_2s^2+L_2s+R_2)(R_3L_3C_3s^2+L_3s+R_3) \\
+R_1L_1(R_2L_2C_2s^2+L_2s+R_2)(R_3L_3C_3s^2+L_3s+R_3)] \\
+\left[\begin{array}{c} L_7L_6C_7C_6s^3+R_6L_7C_6C_7s^2+s[L_7C_7+L_6(C_6+C_7)] \\ +R_6(C_6+C_7)\end{array}\right](C_4C_5s)[[L_7C_7C_6s^2+R_6(C_6+C_7)](R_1L_1C_1s^2+L_1s+R_1) \\
[R_2L_2(R_3L_3C_3s^2+L_3s+R_3)+R_3L_3(R_2L_2C_2s^2+L_2s+R_2)+L_m(R_2L_2C_2s^2+L_2s+R_2)(R_3L_3C_3s^2+L_3s+R_3)]
\end{array}\right]}{[(L_7C_6C_7s^2)+(C_6+C_7)]^3(R_1L_1C_4C_5s^2)\left[\begin{array}{c} R_2L_2(R_3L_3C_3s^2+L_3s+R_3)+R_3L_3(R_2L_2C_2s^2+L_2s+R_2) \\ +L_m(R_2L_2C_2s^2+L_2s+R_2)(R_3L_3C_3s^2+L_3s+R_3)\end{array}\right]}
$$

By substituting the values of all the lumped elements, the resonant frequencies are obtained as 5.467 GHz, 8.382 GHz. Higher frequencies are neglected as they are higher harmonics of the fundamental frequencies. Moreover, the resonant frequencies are mathematically evaluated by neglecting the conductor and dielectric losses. The mathematically computed resonant frequencies are in good agreement with that of the simulated resonant frequencies in Agilent ADS and CST-2016. Figure 6 shows the reflection co-efficient plot of the lumped element equivalent of the proposed patch antenna. Two dips are observed at 5.46 GHz and 8.01 GHz. The lumped element equivalent circuit of the proposed antenna has a good agreement with the simulated results of the proposed patch antenna. A small shift in the frequency is due to the consideration of the lossy inductors and capacitors.

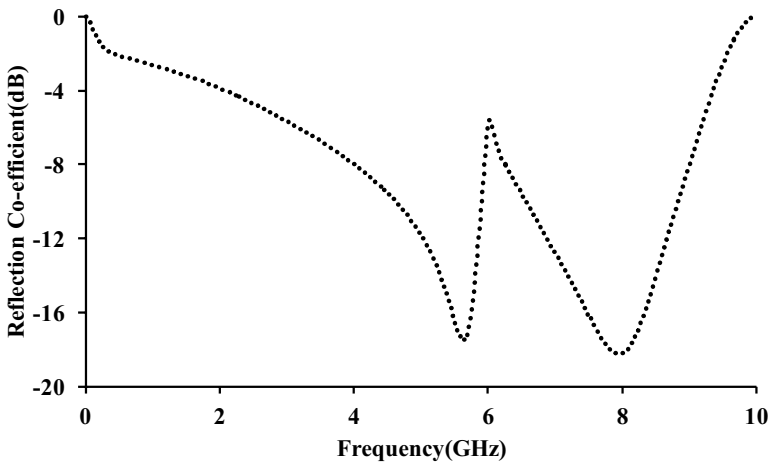


Fig. 6 Prediction of the resonant frequency of the equivalent circuit in Agilent ADS

3 Mathematical Analysis

In order to determine the fields produced by the flower patch antenna with the thickness of 't' μm, Green's function has to be known. Green's functions are originated from the Laplace equations. Basically, Laplace equations determine the shape of the surface. The shape of the proposed flower patch antenna is a series of circles. Hence, the Green's Functions are determined in cylindrical form.

Let G (ρ) be the Green's function.

$$\nabla^2 G = \frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho \frac{\partial G}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 G}{\partial \theta^2} + \frac{\partial^2 G}{\partial z^2} \tag{5}$$

$$\nabla^2 G(\rho) + k_c^2 G\left(\rho, r_o\right) = \frac{1}{\rho}\delta(\rho)\delta(\theta)\delta(z) \tag{6}$$

$$G\ (\rho) = \frac{1}{8\pi^3}\int_0^{k_x}\int_0^{k_y}\int_0^{k_z} \tilde{g}(k) e^{ik.\rho} dk_x dk_y dk_z$$

Where G (ρ) and $\tilde{g}(k)$ form Fourier transform pairs

$$\nabla^2 G\ (\rho) = \frac{1}{4\pi^2 t}\iint_0^{\rho,\psi} \tilde{g}(\rho)\left(-\rho^2\right) e^{i\left(\rho\cos\psi\rho\cos\theta + \rho\sin\psi r'\sin\theta\right)} \rho d\rho d\psi$$

The total transform converts to Henkel Transform or Fourier–Bessel Transform

$$\nabla^2 G\ (\rho) + k_c^2 G\ (\rho) = \frac{1}{4\pi^2 t}\iint_0^{r',\psi} \tilde{g}\left(r'\right)\left(-r'^2\right) e^{i\rho r'\cos(\psi-\theta)} r' dr' d\psi + \frac{k_c^2}{4\pi^2 t}\iint_0^{\rho,\psi} \tilde{g}(\rho) e^{i\rho r'\cos(\psi-\theta)} r' dr' d\psi \tag{7}$$

$$\tilde{g}\left(r'\right) = \int_0^{\rho,}\int_0^{\theta,}\int_0^{z} \frac{1}{\rho}\delta(\rho)\delta(\theta) e^{-i\rho r'\cos(\psi-\theta)} r' dr' d\theta dz = 1 \tag{8}$$

$$\tilde{g}(r') = \frac{t}{k_c^2 - r'^2}$$

Therefore,

$$G(\rho) = \frac{1}{4\pi^2} \iint_0^{r',\psi} \frac{1}{k_c^2 - r'^2} e^{i\rho r' \cos(\psi - \theta)} r' dr' d\psi$$

$$G(\rho) = \frac{1}{2\pi} \int_0^{\infty} \frac{1}{k_c^2 - r'^2} J_o(\rho r') r' dr' = \frac{1}{4\pi k_c} \int_0^{\infty} [J_o(\rho r')] \left[\frac{1}{k_c - r'} + \frac{1}{k_c + r'} \right] r' dr' \quad (9)$$

$$G(\rho) = \frac{1}{4\pi\rho k_c} [\cos(\rho k_c) + \sin(\rho k_c)] \quad (10)$$

$$G(\rho) = \sum_{n=-\infty}^{\infty} J_n\left(\frac{p_{kl}}{r} \rho\right) \sum_{m=-\infty}^{\infty} A_m + \sum_{n=-\infty}^{\infty} J_n\left(\frac{p_{kl}}{r_1 + r_2} \rho\right) \sum_{m=-\infty}^{\infty} A_m \quad (11)$$

$$A_m = \frac{1}{\sum_{m=-\infty}^{\infty} J'_m\left(\frac{p_{kl}}{r} \rho\right)} + \frac{1}{\sum_{m=-\infty}^{\infty} J'_m\left(\frac{p_{kl}}{r_1 + r_2} \rho\right)} \quad (12)$$

$$G(\rho) = \sum_{n=-\infty}^{\infty} \left[J_n\left(\frac{p_{kl}}{r} \rho\right) + J_n\left(\frac{p_{kl}}{r_1 + r_2} \rho\right) \right] \left[\frac{1}{\sum_{m=-\infty}^{\infty} J'_m\left(\frac{p_{kl}}{r} \rho\right)} + \frac{1}{\sum_{m=-\infty}^{\infty} J'_m\left(\frac{p_{kl}}{r_1 + r_2} \rho\right)} \right] \quad (13)$$

where r, r_1 and r_2 are the radius of the centre and edge circular sections of the proposed patch.

The whole patch is analyzed using a cavity model by considering later as a rectangular resonator and the mathematical analysis is carried out by considering proposed as a series of circular waveguides with varying diameters. The metallic patch forms a resonant cavity [22]. Top, bottom, and sides of the cavity are formed by the patch, ground and edges of the patch respectively. The edges of the patch act approximately as an open-circuit boundary. When the patch is excited, the charge is evenly distributed on the upper and lower surfaces of the patch as well as on the ground plane. Hence, the patch acts approximately as a cavity with tangential electric and magnetic fields absent on the top, bottom, and sides of the cavity. The electric field is directed towards z-axis i.e. $\bar{E} = z\,\overline{E_z}$ (x, y). The z-component of electric field satisfies Helmholtz's equation:

$$(\nabla^2 + \beta^2)\bar{E} = 0. \quad (14)$$

where $\beta = \omega\sqrt{\mu\varepsilon}$ is a propagation constant inside the material. $\bar{E}$ is the electric field along z-axis and ∇^2 is the Laplacian operator.

$$\bar{E} = E_o J_n(\rho)\cos(n\phi)$$

$$H_{\phi} = \frac{-jk}{\omega\mu_o} J_n'(\rho)\cos(n\phi)$$

where 'k' is the wave number and $\phi : 0 \rightarrow 2\pi$ and applying the boundary conditions on magnetic field H_{ϕ}.

$H_{\phi}|_{r\,\&(r_1+r_2)} = 0$. The roots of the equations represent different modes in the proposed antenna. From the equivalence principle, the surface currents are given by

$$\bar{M} = 2\overline{E_z}\hat{\phi}$$

Electric vector potential ($\bar{F}$) is given by:

$$\bar{F} = \frac{\varepsilon}{4\pi}\frac{e^{-jk_o r}}{r}\iint \bar{M} e^{jk_o r.\bar{r}} dS$$

'$\bar{r}$' is the source point position and 'k_o' is the wave number in space.

By using the above equations, the electric fields are obtained as:

$$E_{\theta} = \frac{\varepsilon}{4\pi}\frac{e^{-jk_o r}}{r}\sum_{n=-\infty}^{\infty}\left[J_n\left(\frac{p_{kl}}{}r\rho\right) + J_n\left(\frac{p_{kl}}{r_1+r_2}\rho\right)\right]\left[\frac{1}{\sum_{m=-\infty}^{\infty} J_m'\left(\frac{p_{kl}}{}r\rho\right)} + \frac{1}{\sum_{m=-\infty}^{\infty} J_m'\left(\frac{p_{kl}}{r_1+r_2}\rho\right)}\right] \quad (15)$$
$$\left[\left[J_{n+1}(k_o r\sin\theta) - J_{n-1}(k_o r\sin\theta)\right] + \left[J_{n+1}(k_o(r_1+r_2)\sin\theta) - J_{n-1}(k_o(r_1+r_2)\sin\theta)\right]\right]\cos(n\phi)$$

$$E_{\phi} = \frac{\varepsilon}{4\pi}\frac{e^{-jk_o r}}{r}\sum_{n=-\infty}^{\infty}\left[J_n\left(\frac{p_{kl}}{}r\rho\right) + J_n\left(\frac{p_{kl}}{r_1+r_2}\rho\right)\right]\left[\frac{1}{\sum_{m=-\infty}^{\infty} J_m'\left(\frac{p_{kl}}{}r\rho\right)} + \frac{1}{\sum_{m=-\infty}^{\infty} J_m'\left(\frac{p_{kl}}{r_1+r_2}\rho\right)}\right] \quad (16)$$
$$\left[\left[J_{n+1}(k_o r\sin\theta) - J_{n-1}(k_o r\sin\theta)\right] + \left[J_{n+1}(k_o(r_1+r_2)\sin\theta) - J_{n-1}(k_o(r_1+r_2)\sin\theta)\right]\right]\cos\theta\sin(n\phi)$$

Figures 4, 5 and 6 show the plot of mathematically evaluated Green's function of the proposed patch antenna as obtained in Eq. (13).

Clearly, the plot in Fig. 7 shows the maximization of the Green's function at lower distances and it reaches a minimum value at larger distances. The radiation pattern of an antenna depends on the current distribution and the Green's function.

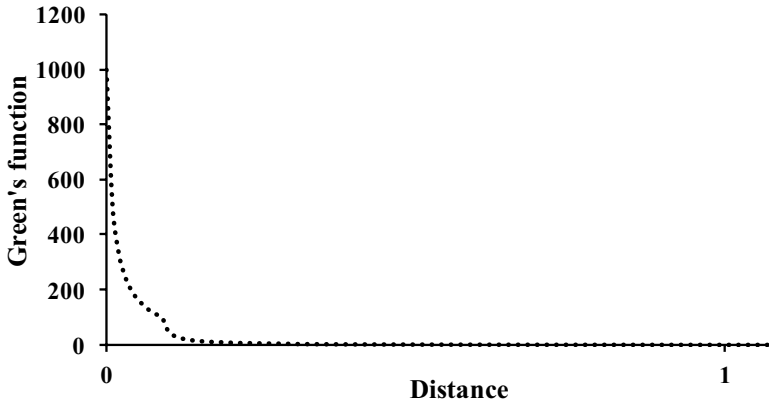


Fig. 7 Plotted Green's function of the proposed patch antenna

4 Parametric Analysis

The performance of the proposed antennas is decided by the various structural parameters. The effects of the key features like the variation in the gain and reflection coefficients with the variation in the dimension of the Jerusalem cross-shaped DGS, number of metamaterial cells, and height of the foam as observed through CST simulations are discussed in this section. Figure 8 shows the simulated return loss for 9 metamaterial cells, 1 metamaterial cell and without metamaterial cells. As the metamaterial cells decrease, the resonant frequencies shift towards the lower side and upper side of the desired frequency ranges with the degradation in the gain. When 9 cells are embedded, the proposed patch antenna resonates at 3.2 GHz, 4.01 GHz and 8.1 GHz respectively. The bandwidths observed at these frequencies are 101Mz, 189 MHz and 301 MHz respectively. The corresponding gains at these frequencies are 2.1 dBi, 3.21 dBi and 4.8 dBi as shown in Fig. 11 (Dashed green line). When 1 metamaterial cell is embedded, two resonant frequencies namely 3.85 GHz and 8.02 GHz are observed with the bandwidths of 262 MHz and 980 MHz respectively. The corresponding gains at these frequencies are 3.2 dBi and 2.6 dBi as shown in Fig. 11 (Dashed red line). The proposed patch without metamaterials resonates at 7.8 GHz and 9.56 GHz respectively. The bandwidths observed at these frequencies are 78 MHz and 52 MHz respectively. The corresponding gains at these frequencies are 1.1 dBi and 2.01 dBi as shown in Fig. 11 (Blue dotted line).

The variation in the foam substrate height makes the patch antenna resonate at multiple frequencies with the decrement in the bandwidth. Figure 9 shows the variation in the height of the foam substrate for 25 metamaterial cells. A foam thickness of 1 mm makes the proposed patch antenna resonate at 4.25 GHz, 4.52 GHz and 8.15 GHz respectively. The bandwidths at observed at 4.25 GHz and 8.15 GHz are 1.1 GHz and 612 MHz respectively. When the foam thickness is reduced to 0.5 mm, patch resonates at 2.52 GHz, 6.72 GHz and 8.12 GHz respectively. The bandwidths observed at these frequencies are 115 MHz, 560 MHz and 430 MHz respectively. Further, the proposed patch without the foam substrate resonates at 2.61 GHz, 7.72 GHz and 9.81 GHz respectively. The bandwidths observed at these frequencies are 120 MHz, 542 MHz and 631 MHz respectively. Moreover, impedance matching is highly affected by the decrease in the height of the foam substrate. Increasing the height of the foam substrate increases the physical dimensions of the antenna. An increase of foam width beyond 2 mm will produce unacceptable results and hence, they are not reported in this paper. Hence, the parametric analysis on the thickness of the foam substrate is not carried out beyond 2 mm.

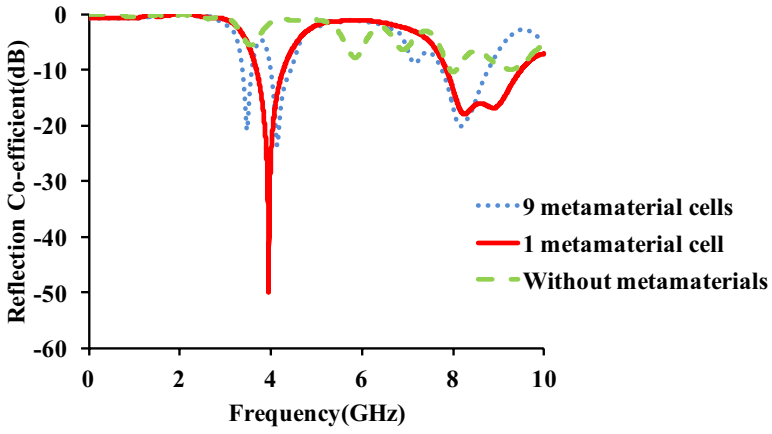


Fig. 8 Effect of the metamaterial cells on the return loss of the flower patch antenna

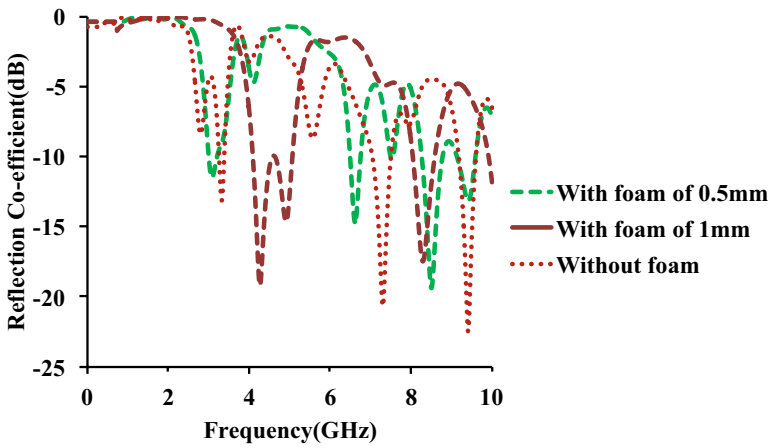


Fig. 9 Effect of the foam thickness on the return loss of the flower patch antenna

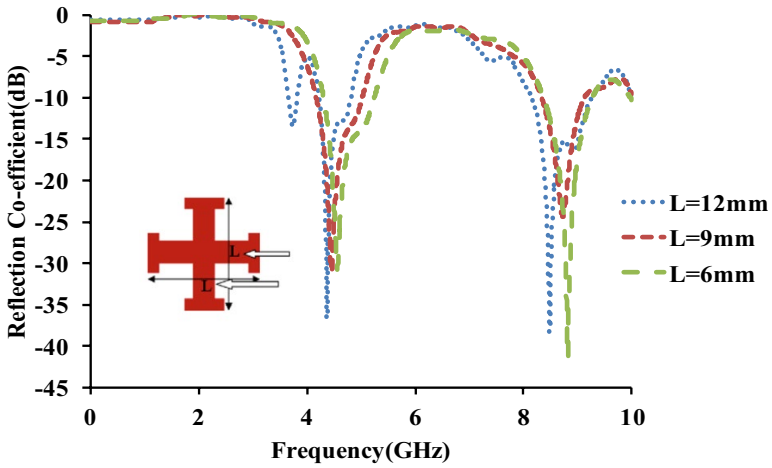


Fig. 10 Effect of the dimensions of the Jerusalem cross-shaped DGS on the return loss of the flower patch antenna

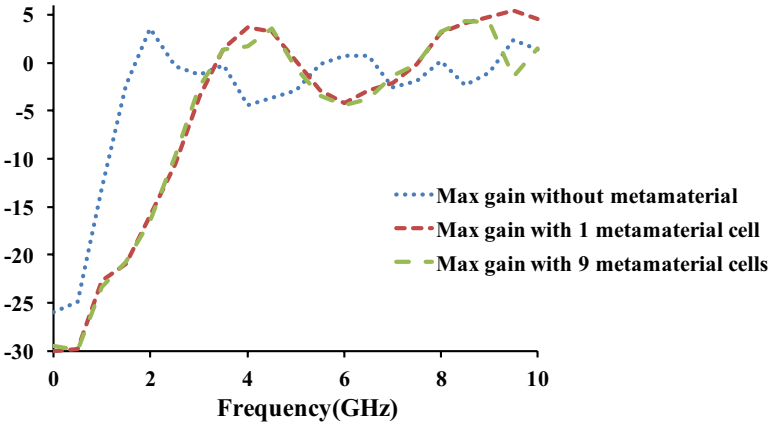


Fig. 11 Effect of the metamaterial cells on the gain of the flower patch antenna

Varying the dimensions of the Jerusalem cross-shaped DGS has little impact on the change in the resonant frequencies of the patch. But, there is degradation in the bandwidth at the desired resonant frequencies of the patch antenna. Figure 10 shows the variation in the dimensions of the Jerusalem cross-shaped DGS and its impact on the proposed patch's performance. By choosing 'L' = 12 mm, the patch antenna resonates at three frequencies namely 2.67 GHz, 4.57 GHz and 8.36 GHz with the degradation in the bandwidths. If the 'L' is further decreased to 9 mm then antenna resonates at 4.62 GHz and 8.43 GHz respectively. Bandwidths of 1.01 GHz and 762 GHz are obtained in both the bands. Further

decrease in the dimensions (L = 6 mm) of the Jerusalem cross-shaped DGS has little impact on the frequency of operation and the bandwidth.

Hence, an optimization is achieved on the number of metamaterial cells, dimensions of the Jerusalem cross-shaped DGS and thickness of the foam in order to obtain a trade-off between gain and bandwidth of the flower patch antenna. Increasing the dimensions of the Jerusalem cross increases the gain of the proposed patch antenna but with the degradation in the impedance bandwidth. Further increase in the dimensions of the Jerusalem cross-shaped DGS would reduce the bandwidth along with the degradation in the gain and the radiation pattern. Figure 11 shows the change in the gain with the number of metamaterial cells. Gains of − 3.2 dBi and 1.2 dBi are observed at 5.2 GHz and 8.25 GHz with 9 metamaterial cells as represented in green dashed line. Gains of − 2.8 dBi and 1.2 dBi are observed at 5.2 GHz and 8.25 GHz with one metamaterial cell as represented in red dashed line. Further, gains of − 5 dBi and − 2.1 dBi are observed at 5.2 GHz and 8.25 GHz without metamaterial cell as shown in blue dotted line.

Figure 12 shows the effect of rotation of the DGS on the reflection co-efficient of the flower patch antenna. When the Jerusalem cross-shaped DGS is rotated by 15 degrees, the proposed antenna resonates at 2.56 GHz, 6.06 GHz, 8.65 GHz and 9.81 GHz. The impedance bandwidth reduces to 210 MHz, 171 MHz, 105 MHz and 436 MHz in the frequency bands of 2.49–2.7 GHz, 6.01–6.271 GHz, 8.61–8.72 GHz and 9.58–10 GHz respectively. The proposed flower patch antenna resonates at 3.71 GHz, 4.21 GHz, 7.34 GHz and 7.85 GHz when the Jerusalem DGS is rotated by 30 degrees. The impedance bandwidths of 101 MHz, 123 MHz, and 724 MHz are observed in the frequency bands of 3.7–3.8 GHz, 4.15–4.27 GHz, 7.27–8 GHz respectively. If the DGS is rotated by 45 degrees then the proposed patch resonates at 2.56 GHz and 6.21 GHz. The impedance bandwidth reduces to 210 MHz and 322 MHz in the frequency ranges of 2.49–2.7 GHz and 6.01–6.321 GHz respectively. The rotation of the DGS by 60 degrees makes the proposed antenna resonate at 2.56 GHz, 6.39 GHz and 8.41 GHz with impedance bandwidths of 210 MHz, 133 MHz and 103 MHz respectively. Further rotation of the DGS yields unacceptable results and hence those results are not reported in this paper.

Figure 13 shows the placement of the DGS at different points on the ground plane. The displacement of DGS by 3.75 mm from the origin in the negative x-direction resonates the antenna at 6.89 GHz, 8 GHz and 8.21 GHz with the impedance bandwidths of 81 MHz, 101 MHz and 93 MHz respectively. The displacement of the DGS by 3.75 mm from the

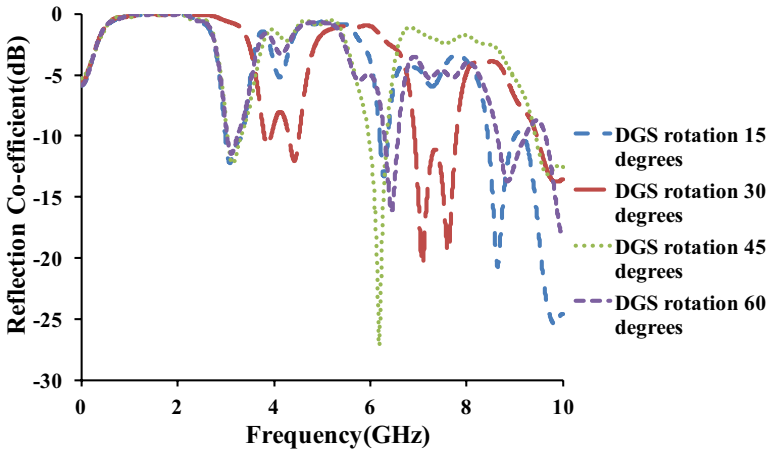


Fig. 12 Effect of rotation of the DGS on the impedance bandwidth of the flower patch antenna

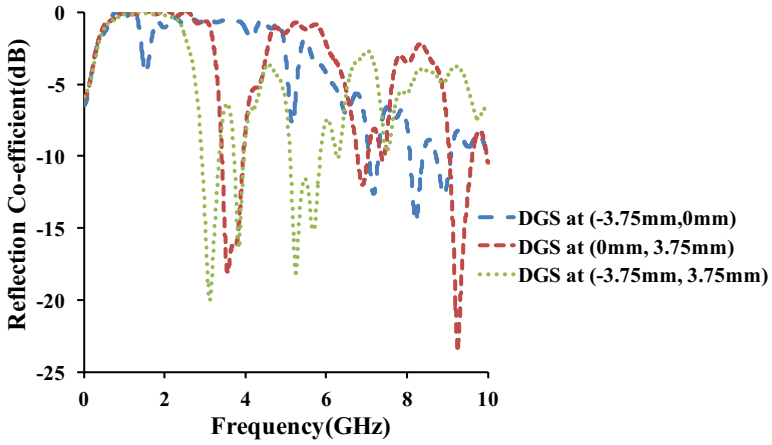


Fig. 13 Effect of the position of the DGS on the impedance bandwidth of the flower patch antenna

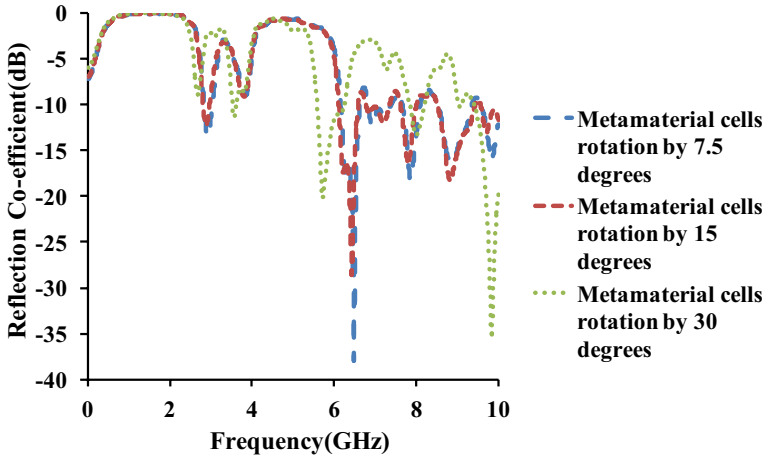


Fig. 14 Effect of the rotation of the metamaterial cells on the impedance bandwidth of the flower patch antenna

origin in the positive y-direction makes the antenna resonate at 3.75 GHz, 6.67 GHz, 7.23 GHz and 9.71 GHz. The impedance bandwidths are observed to be 534 MHz, 111 MHz, 85 MHz and 223 MHz respectively. Further, the DGS is placed at 3.75 mm from the origin in the negative x-axis and 3.75 mm from origin in the positive y-axis. The proposed patch resonates at 2.58 GHz, 5.25 GHz, 5.76 GHz with the impedance bandwiths of 336 MHz and 762 MHz respectively.

Figure 14 shows the variation of the impedance bandwidth with the metamaterial cell rotations. The rotation of the metamaterial cells by 7.5 degrees makes the patch resonate at 2.56 GHz, 6.21 GHz, 6.85 GHz, 7.89 GHz and 8.36 GHz. The impedance bandwidths are observed to be 102 MHz, 223 MHz, 65 MHz and 156 MHz respectively. If the metamaterial cells are rotated by 15 degrees then the patch resonates at the same frequencies as that of the previous case with the improvement in the return losses. Further rotation of the metamaterial cells by 30 degrees shifts the resonant frequencies to 3.52 GHz, 5.82 GHz, 8.1 GHz and 9.89 GHz. The impedance bandwidths in these bands are observed to be 105 MHz, 335 MHz and 201 MHz respectively

Figure 15 shows the change in the impedance bandwidths of the proposed patch antenna with the dimensions of the metamaterial cells. The patch antenna resonates at 2.23 GHz, 3.12 GHz, 3.68 GHz, 6.21 GHz, 7.9 GHz, 9.3 GHz and 9.91 GHz with the metamaterial cell area of 16 mm^2. The impedance bandwidths of 106 MHz, 666 MHz, 332 MHz, 501 MHz and 725 MHz are observed in the respective bands. The patch antenna resonates at 2.56 GHz, 3.34 GHz, 6.21 GHz, 7.83 GHz, 8.1 GHz and 9.34 GHz with the

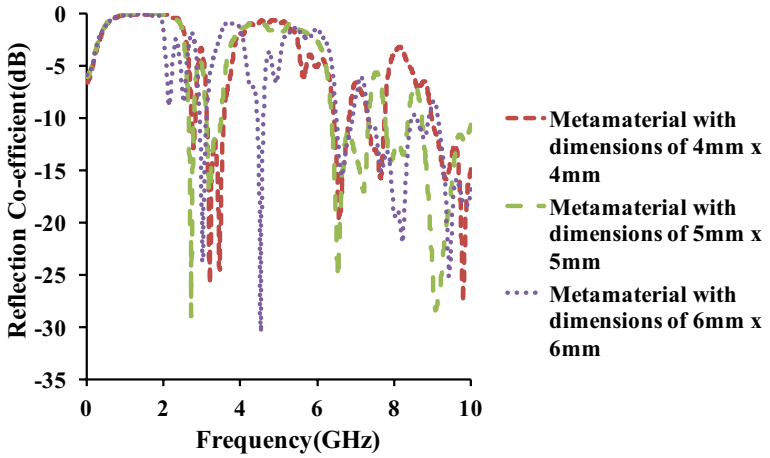


Fig. 15 Effect of the dimensions of the metamaterial cells on the impedance bandwidth of the flower patch antenna

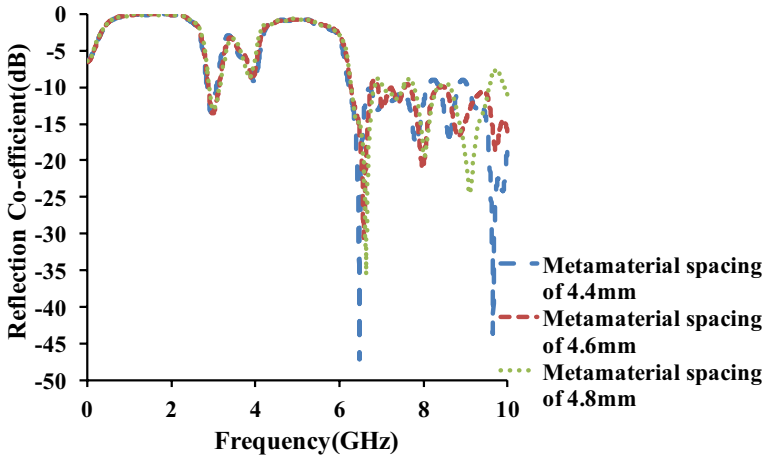


Fig. 16 Effect of the spacing between the metamaterial cells on the impedance bandwidth of the flower patch antenna

metamaterial cell area of 25 mm^2. The impedance bandwidths of 106 MHz, 256 MHz, 782 MHz, 565 MHz and 725 MHz are observed in the respective bands. When the area of the metamaterial cell is further increased to 36 mm^2, the patch resonates at 2.73 GHz, 4.34 GHz, 6.21 GHz, 8.09 GHz and 9.75 GHz. The impedance bandwidths of 106 MHz, 256 MHz, 782 MHz, 565 MHz and 725 MHz are observed in the respective bands.

Figure 16 shows the variation of the impedance bandwidth with the spacing between the two metamaterial cells. A spacing of 4.4 mm makes the patch antenna resonate at 2.76 GHz, 6.25 GHz, 7.86 GHz, 8.38 GHz and 9.83 GHz. The impedance bandwidths in these bands are 212 MHz, 178 MHz, 212 MHz, 99 MHz and 158 MHz respectively. With the metamaterial cell spacing of 4.6 mm, the proposed patch resonates at 2.76 GHz, 6.25 GHz, 7.75 GHz, 8.31 GHz and 9.83 GHz. The impedance bandwidths in these bands are 156 MHz and 3.46 GHz respectively. By increasing the metamaterial cell spacing by 4.8 mm, the proposed patch resonates at 2.76 GHz, 6.25 GHz, 7.75 GHz, 8.31 GHz and 9.56 GHz. The impedance bandwidths in these bands are observed to be 156 MHz, 103 MHz, 201 MHz and 436 MHz respectively.

5 Results and Discussions

The optimized dimensions of Jerusalem cross-shaped DGS are obtained as 15 mm × 15 mm. By introducing the foam of 2 mm, the antenna gives best results.

Figure 17 shows the simulated and measured return losses of the proposed patch antenna. The simulated bandwidth of 23% in the range of 4.95–6.15 GHz and 26.9% in the range of 7.1–9.3 GHz is obtained. Measured bandwidths of 22.8% and 25.7% are obtained in the bands of 4.95–6.15 GHz and 7.1–9.3 GHz respectively. The measured and simulated results are in good agreement. Radiation efficiencies of 61% and 69.7% are obtained at 5.25 GHz and 8.25 GHz as shown in Fig. 18.

The radiation patterns of the proposed patch antenna are simulated in E-plane and H-plane using CST Microwave Studio and measured in an anechoic chamber using an antenna measurement system. The reference antenna used for testing is a double-ridged horn. Figure 19 shows the measured and simulated results of the gain in E &H planes. A simulated gain of 3.93 dBi is obtained in both E ($\varphi = 90^{\circ}$) and H ($\varphi = 0^{\circ}$) planes at 5.2 GHz. The measured values of the gain at the same frequency are 3.75 dBi and 3.8 dBi respectively. Similarly, simulated gains of 4.92 dBi and 5.03 dBi are obtained in E ($\varphi = 90^{\circ}$) and H ($\varphi = 0^{\circ}$) planes. The measured values of the gain at the same frequency are obtained as 5.02 dBi and 4.56 dBi respectively. It can be seen that the simulated and measured results are in good agreement.

Figure 20 shows the simulated and measured gains of the proposed patch antenna. A measured gain of 3.75 dBi is obtained at 5.2 GHz and 5.02 dBi at 8.25 GHz. For calculating the permittivity and permeability, Nicolson-Ross-Weir is used. The above method is carried out by extracting s-parameters from the CST-MWS package and the subsequent calculations are done using MATLAB. The magnitude and phase of S_{11} and S_{21} for the unit cell are extracted using CST-MWS package as shown in Fig. 21a, b, and the permittivity and permeability are obtained by MATLAB [23].

From the Fig. 22a, it can be seen that the triple CSRR metamaterial exhibits negative permittivity (dark line) and permeability (dotted line) at 5.2 GHz and 8.25 GHz respectively. It can be concluded that the triple CSRR metamaterial is a Double Negative Metamaterial (DNG MTM). It can further be observed that the metamaterial possesses a near

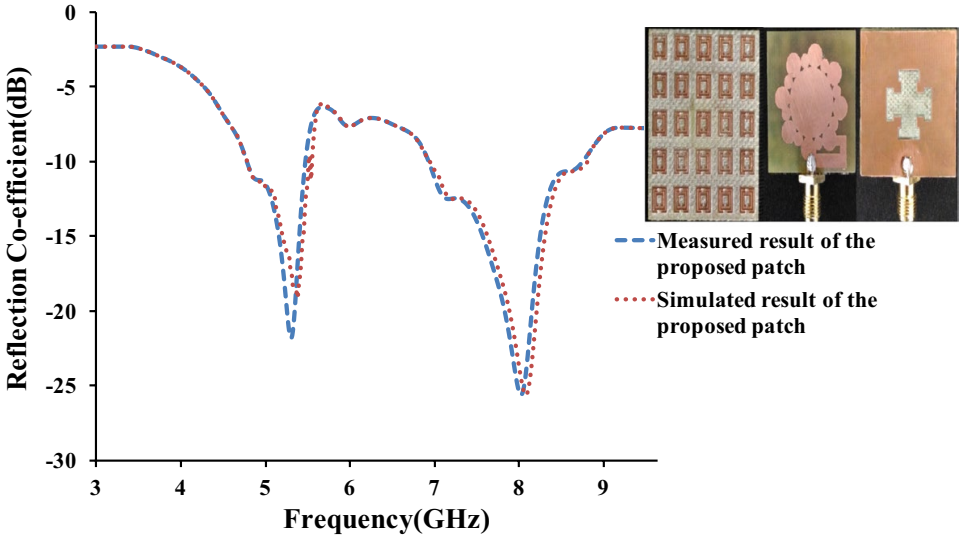


Fig. 17 Simulated and measured results of the reflection coefficient of the proposed antenna

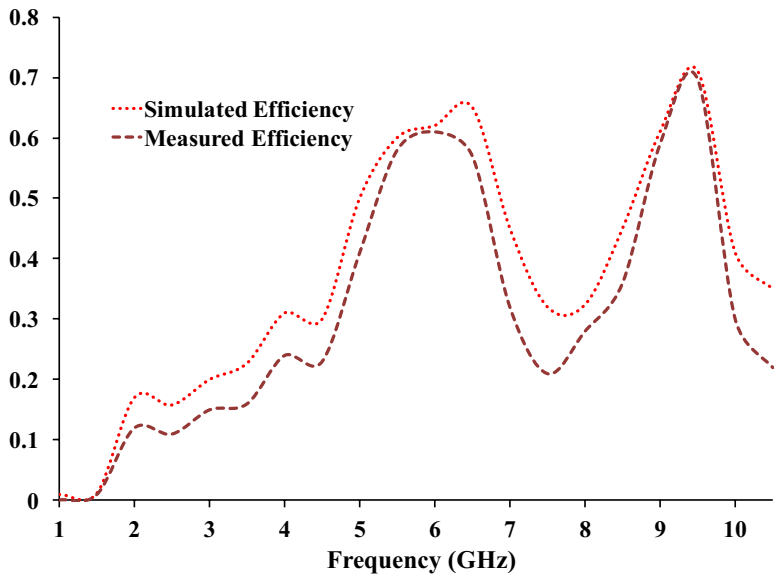


Fig. 18 Simulated and measured efficiency of the proposed antenna

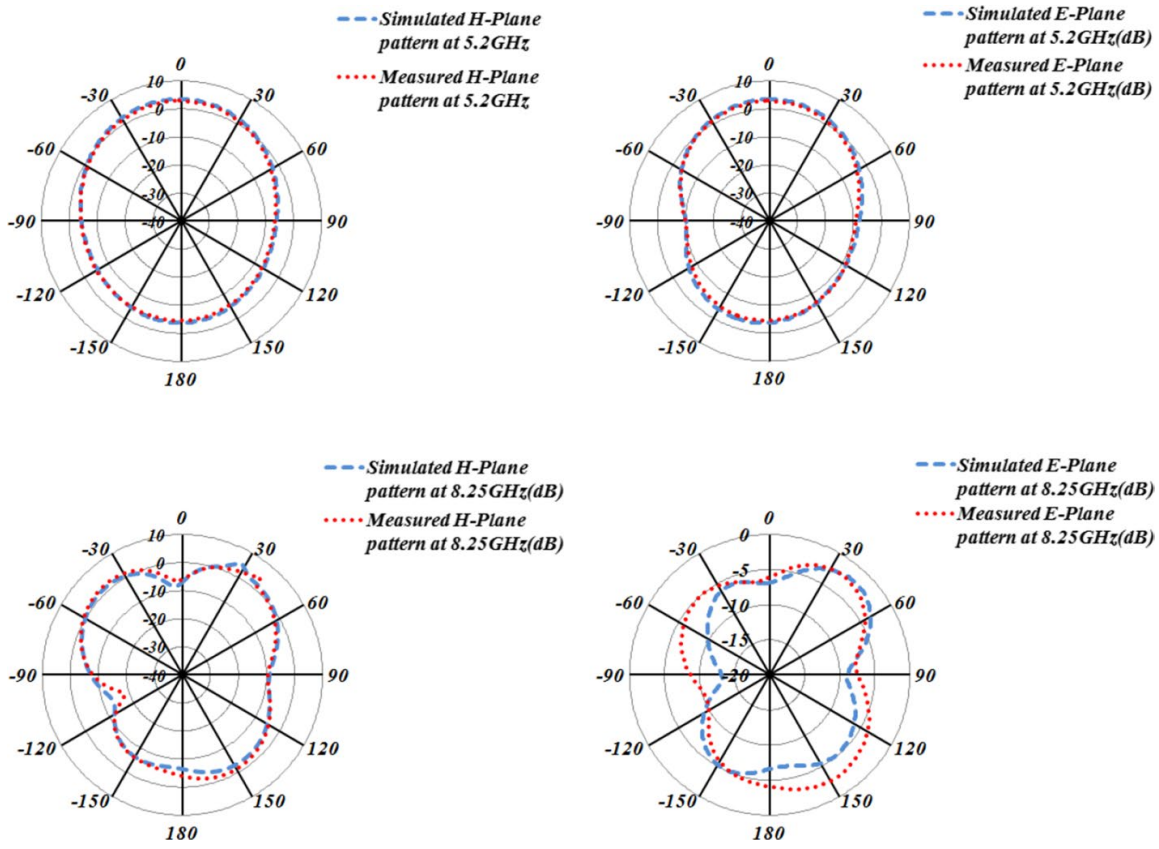


Fig. 19 Simulated and measured gain of the proposed patch antenna

periodic negative permittivity and permeability in the window of 5 GHz- 9 GHz. Figure 22b shows the plots of imaginary permittivity (in thick line) and permeability (in dotted line). Plots of the negative refractive index and the impedance are indicated with thick dark line and the dotted as shown in Fig. 22c. Figure 22d shows the plots of the imaginary refractive index and the impedance of the triple CSRR metamaterial. Figure 23 shows the surface current density of antenna at 5.2 GHz and 8.25 GHz respectively. From the Fig. 23, it can be concluded that the resonance at 5.2 GHz and 8.25 GHz is due to metamaterial and

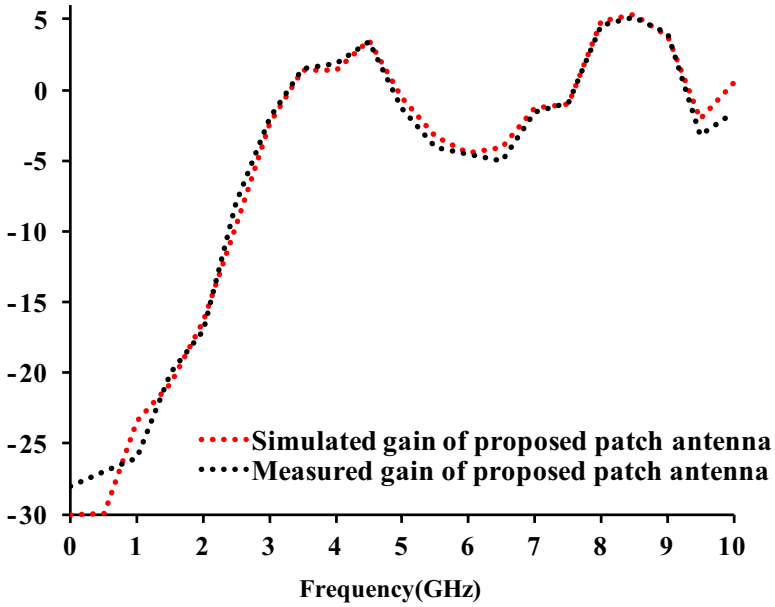


Fig. 20 Simulated and measured gains of the proposed patch antenna

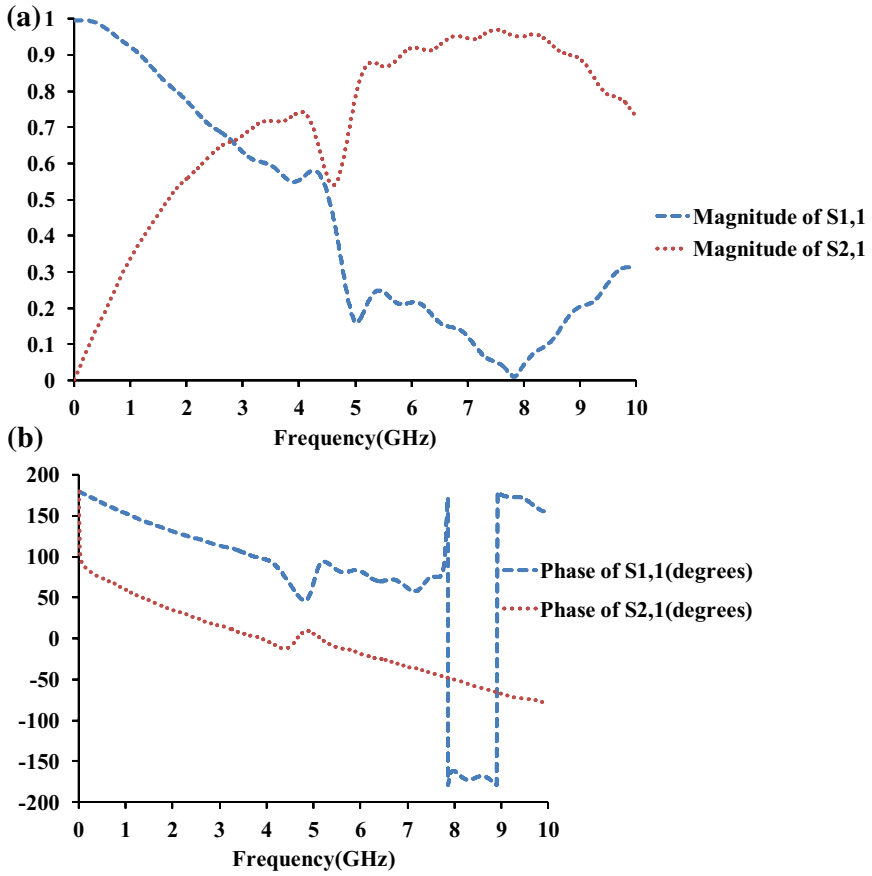


Fig. 21 **a** Magnitude plots of the return and insertion losses of a triple CSRR unit cell. **b** Phase plot of the return and insertion losses of a triple CSRR unit cell

the Jerusalem shaped DGS ground plane as they possess maxima along their lengths and widths respectively. The maximum radiation is observed along the edges of the metamaterial cells.

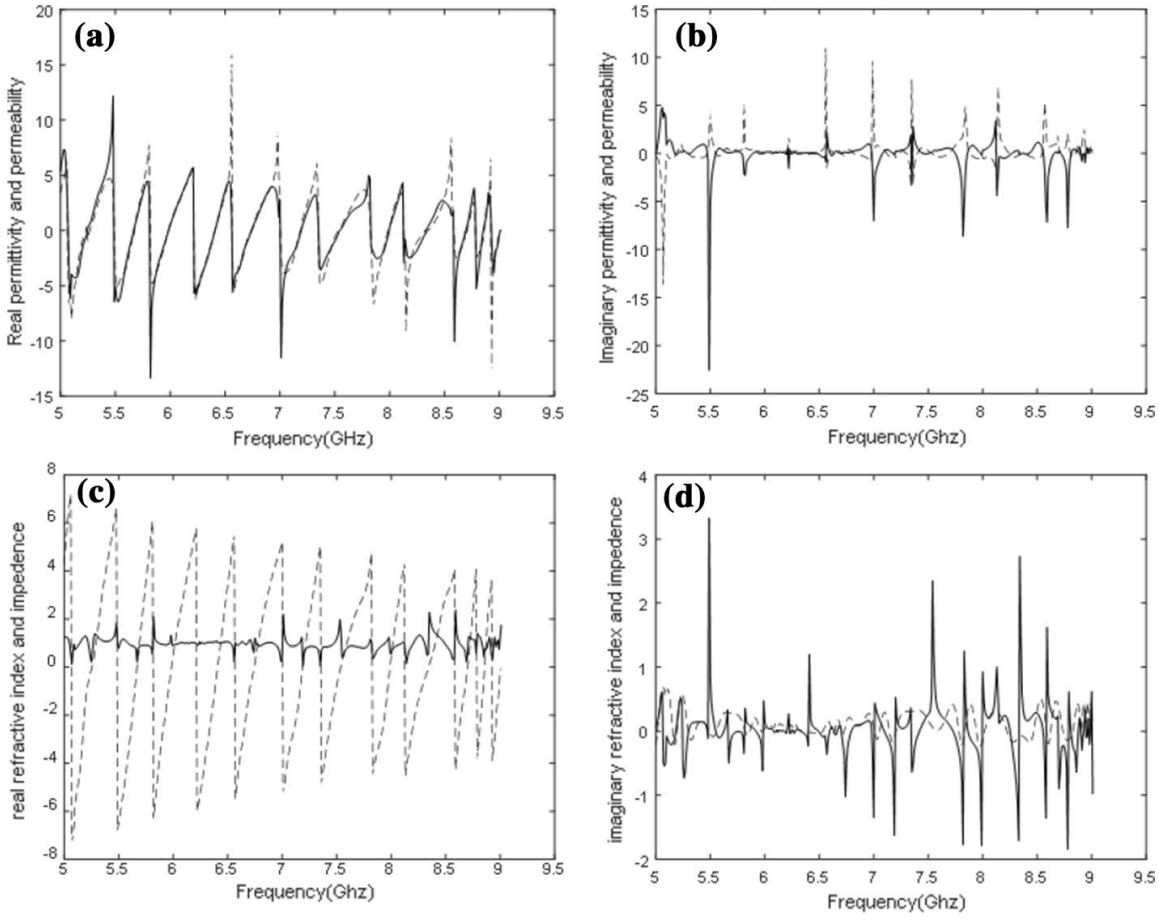


Fig. 22 **a** The real part of permittivity and permeability. **b** The imaginary part of permittivity and permeability. **c** The real part of refractive index and impedance. **d** Imaginary part of refractive index and impedance

Table 3 shows the comparison of the proposed patch antenna with existing antennas. The proposed antenna possesses wide bandwidth in both the bands compared to the

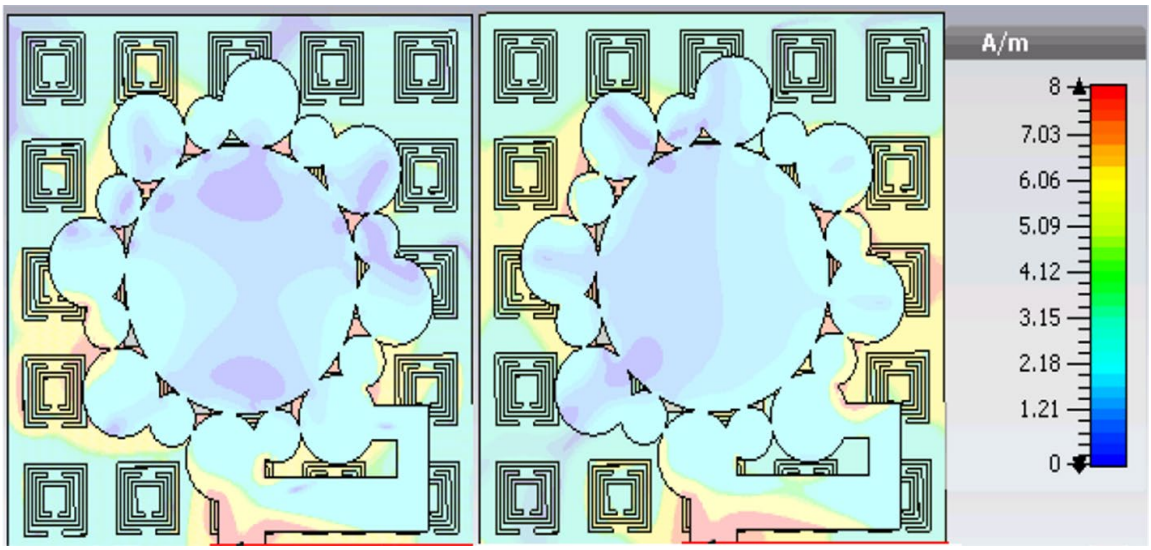


Fig. 23 Surface current distribution in the proposed patch antenna

Table 3 Comparison between the proposed antenna and existing antennas

Reference	Dimensions of the patch (mm^2)	Area (mm^2)	Centre frequency (GHz)	Bandwidth (GHz)	Gain (dBi)	Application (s)
Guo et al.	30×30	900	2.4/5.5	0.08/0.51	Not mentioned	Not mentioned
Shukla et al.	70×50	3500	1.71/2.28/3.03/3.84	2.78	2.85/2.98/4.04/4.73	L- and S-band applications
Kaushal et al.	60×60	3600	3.18/4.31/6.42/7.58/11.28/12.64	0.108/0.10/0.142/0.112/0.008/0.036	2.32/1.25/3.89/4.1/6.74/7.2	Space and military applications
Khandelwal et al.	46.4×46.4	2152.96	2.55/5.25	Not mentioned	6/7.2	WiMAX/WLAN
Awais et al.	25×35	875	2.45/3.55	1.6/0.5	8.9	Bluetooth/IOT/5G
Kumar et al.	32.7×39.5	1291.65	2.4/2.6	Not mentioned	3.65/4.71	Bluetooth
Saroj et al.	46×44	2024	1.07/3.119/4.089/5.683/6.514	Not mentioned	Not mentioned	L-/S-/C-band applications
Kaur et al.	48×41	1968	2.5/3.51	0.25/0.37	4.08/5.024	WLAN/Bluetooth/ZigBee/Wi-Max/IMT band applications
Khanna et al.	25×25	625	3.23/4.93/7.09	4.99	1.1/0.8/1.02	S-/C-band applications
Ali et al.	30×24.8	744	3.3/5.5/7.3/9.9	0.11/0.28/1.01/0.56	0.7/0.65/1.1/1.7	WiMAX, and X-band applications
Bhatia et al.	17×23	391	4.4/6.8/9.6	4.35	7.5	WiMAX, and X-band applications
Goswami et al.	50.9×33	1679.7	2.9/3.4/7.5	0.2/4.2	2.1/1.3	UWB applications
Proposed antenna	23.5×16	376	5.25/8.25	1.2/2.2	3.75/5.2	WLAN/WiMAX

different antennas. The proposed antenna occupies less area with improved bandwidth and gain than the existing antennas.

6 Conclusion

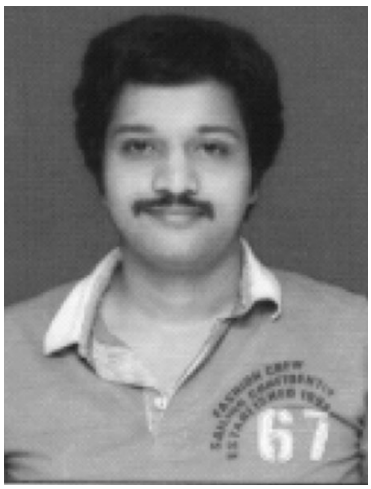
In this paper, a novel and dual wideband patch antenna is designed and analysed using CST-MWS package. The present study infers that the gain and bandwidth are controlled by introducing foam and double feeding the patch. The effective size of the antenna is significantly reduced with the maximum bandwidth of 26.50% by introducing triple CSRR metamaterials and Jerusalem cross-shaped DGS. Therefore, the proposed antenna is desirable for WLAN/WiMAX applications.

References

1. Raghava, N. S., & De, A. (2006). Photonic band gap stacked rectangular microstrip antenna for road vehicle communication. *IEEE Transactions of Antenna and Wireless Propagation Letters, 5,* 4241–4423.
2. Raghava, N. S., & De, A. (2007). A novel high performance patch radiator. *International Journal in Microwave Science and Technology, 2008,* 1–4.
3. Raghava, N. S., & De, A. (2009). Effect of an air gap width on the performance of a stacked square electronic band gap antenna. *International Journal of Microwave and Optical Technology, 4*(5), 315–317.
4. Zakaria, N. A., Sulaiman, A. A., & Latip, M. A. A. (2008). Design of a circular microstrip antenna. In *IEEE-international RF and microwave proceedings* (pp. 289–292), Malaysia.
5. Paramayudha, K., Santiko, A. B., Wahyu, Y., Oktafiani, F., Fitriadi, A., & Wijanto, H. (2016). Design and realization of circular patch antenna for S-Band Coastal Radar. In *International conference on radar, antenna, microwave, electronics, and telecommunications (ICRAMET)* (pp. 115–118), Indonesia.
6. Saleem, R., & Quddus, A. (2017). Circular slotted reconfigurable multiband patch antenna for wireless applications. In *International symposium on applied computational electromagnetics society (ACES)* (pp. 1–2), Italy.
7. Guo, X., Liao, W., Zhang, Q., & Chen, Y. (2016). A dual-band embedded inverted T-slot circular microstrip patch antenna. In *IEEE 5th Asia-Pacific conference on antennas and propagation (APCAP)* (pp. 151–152), Taiwan.
8. Sharma, S., & Tripathi, C. C. (2017). A novel reconfigurable antenna with spectrum sensing mechanism for CR system. *Progress in Electromagnetics Research C, 72,* 187–196.
9. Liu, Y., Wang, H., Xue, F., & Dong, X. (2017). A new single-layer reflectarray using circular patch with semicircular ring slots. *Progress in Electromagnetics Research Letters, 66,* 105–111.
10. Shukla, B. K., Kashyap, N., & Baghel, R. K. (2017). Circular slotted elliptical patch antenna with elliptical notch in ground. *Progress in Electromagnetics Research C, 74,* 181–189.
11. Kurniawan, F., Sri, J. T., Sumantyo, K. I., Kuze, H., & Gao, S. (2017). Patch antenna using rectangular centre slot and circular ground slot for circularly polarized synthetic aperture radar (CP-SAR) application. *Progress in Electromagnetics Research, 160,* 51–61.
12. Tiwari, R. N., Singh, P., & Kanaujia, B. K. (2017). Butter fly shape compact microstrip antenna for wideband applications. *Progress in Electromagnetics Research Letters, 69,* 45–50.
13. Dan, Yu., Wan, Y.-T., & Zhang, H.-B. (2017). Low-profile wideband circularly polarized microstrip antenna with conical radiation pattern. *Progress in Electromagnetics Research C, 72,* 81–89.
14. Kaushal, D., & Shanmuganantham, T. (2018). Parametric enhancement of a novel microstrip patch antenna using Circular SRR Loaded Fractal Geometry. *Alexandria Engineering Journal, 57,* 2551–2557.
15. KhalidHati, N., ElHajjaji, A., & Asselman, H. (2016). Design of new antenna in the form of dollar-symbole for WLAN technology. *Procedia Technology, 22,* 606–613.
16. Kumar Naik, K., & Amala Vijaya Sri, P. (2018). Design of hexadecagon circular patch antenna with DGS at Ku band for satellite communications. *Progress in Electromagnetics Research M, 63,* 163–173.

17. Dawar, P., Raghava, N. S., & De, A. (2017). Miniaturized UWB multi-resonance patch antenna loaded with novel modified H-shape SRR metamaterial for microspacecraft applications. *Frontiers of Information Technology & Electronic Engineering, 18*(11), 1883–1891.
18. Dawar, P., De, A., & Raghava, N. S. (2016). Directive and broadband 4-seg SRR metamaterial antennas. *International Journal of Advanced Science and Technology, 97,* 1–12.
19. Dawar, P., Raghava, N. S., & De, A. (2015). Tunable and directive metamaterial-inspired antennas for 'C' band applications. *International Journal of Microwave and Optical Technology, 10*(3), 168–175.
20. Dawar, P., Raghava, N. S., & De, A. (2016). High gain, directive and miniaturized metamaterial C-band antenna. *Cogent Physics, 3,* 1236510.
21. Bahl, I., Bhartia, P., Garg, R., & Ittipiboon, A. (2000). *Microstrip antenna design handbook* (2nd ed.). London: Artech House (Antenna and Wave propagation Library).
22. Viswanadha, K., & Nallanthighal, R. (2018). Design and analysis of a meander line cornered microstrip patch antenna with square slotted EBG structure for ISM/WLAN applications. *International Journal of Advanced Science and Technology, 113,* 93–102.
23. Rothwell, E. J., Frasch, J. L., Ellison, S. M., Chahal, P., & Ouedraogo, R. O. (2016). Analysis of the Nicolson–Ross–Weir method for characterizing the electromagnetic properties of engineered materials. *Progress in Electromagnetics Research, 157,* 31–47.
24. Khandelwal, M. K., Kanaujia, B. K., Dwari, S., Kumar, S., & Gautam, A. K. (2015). Analysis and design of dual band compact stacked Microstrip patch antenna with defected ground structure for WLAN/WiMax applications. *AEU - International Journal of Electronics and Communications, 69*(1), 39–47.
25. Awais, Q., Chattha, H. T., Jamil, M., Jin, Y., Tahir, F. A., & Rehman, M. U. (2018). A novel dual ultrawideband CPW-fed printed antenna for internet of things (IoT) applications. *Wireless Communications and Mobile Computing, 2018,* 1–9.
26. Kumar, S., Singh, A. P., & Khandelwal, M. K. (2017). Theoretical analysis and design of dual band DGS antenna with small frequency ratio for Wi-Fi and WiMAX applications. *Progress in Electromagnetics Research M, 62,* 153–166.
27. Saroj, A. K., Mohd, G., Siddiqui, M. K., & Ansari, J. A. (2017). Design of multiband quad-rectangular shaped microstrip antenna for wireless applications. *Progress in Electromagnetics Research M, 59,* 213–221.
28. Kaur, A., Khanna, R., & Kartikeyan, M. (2017). A multilayer dual wideband circularly polarized microstrip antenna with DGS for WLAN/Bluetooth/ZigBee/Wi-Max/IMT band applications. *International Journal of Microwave and Wireless Technologies, 9*(2), 317–325.
29. Khanna, P., Sharma, A., Shinghal, K., & Kumar, A. (2016). A defected structure shaped CPW-fed wideband microstrip antenna for wireless applications. *Journal of Engineering, 2016,* 1–7.
30. Ali, T., Aw, M. S., & Biradar, R. C. (2018). A fractal quad-band antenna loaded with L-shaped slot and metamaterial for wireless applications. *International Journal of Microwave and Wireless Technologies*. https://doi.org/10.1017/S1759078718000272.
31. Bhatia, S. S., Sahni, A., & Rana, S. B. (2018). A novel design of compact monopole antenna with defected ground plane for wideband applications. *Progress in Electromagnetics Research M, 70,* 21–31.
32. Goswami, S. A., & Karia, D. (2017). A compact monopole antenna for wireless applications with enhanced bandwidth. *AEU - International Journal of Electronics and Communications, 72,* 33–39.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Mr. Karteek Viswanadha received B.Tech. in electronics and communication from G.V.P College of Engineering (Autonomous), Visakhapatnam in 2007 and master in communication systems from Birla Institute of Technology and Sciences, Pilani (B.I.T.S, Pilani) in 2009. He is presently pursuing Ph.D. in the area of millimeter wave devices and antennas from Delhi Technological University (Formerly Delhi College of Engineering), New Delhi.

Mr. Nallanthighal Srinivasa Raghava is working as Professor in Electronics and Communication Engineering Department in Delhi Technological University (Formerly Delhi College of Engineering), New Delhi. His area of specialization is Antenna and Propagation, Microwave Engineering, Digital Communication, Wireless Communication, Cloud Computing, Information Security.

Discrete Mathematics and Theoretical Computer Science DMTCS vol. **VOL**:ISS, 2020, #NUM

Domination in Signed Petri Net

Payal[1*] Sangita Kansal [1 †]

[1] *Department of Applied Mathematics,Delhi Technological University*

received 2020-01-14, *revised 2020-01-14*, *accepted 2020-02-14*.

In this paper, domination in Signed Petri net(SPN) has been introduced.We identify some of the Petri net structures where a dominating set can exist.Applications of producer consumer problem, searching of food by bees and finding similarity in research papers are given to understand the areas where the proposed theory can be used.

Keywords: Petri net; Signed graph; Domination in graph

arXiv:2001.04374v1 [cs.DM] 13 Jan 2020

1 Introduction

The foundations of theory of domination can be traced back to the chess problem of finding the minimum number of queens required such that all the squares are either occupied or can be attacked by a queenJaenisch (1862).The applications of theory of domination includes communication network problems, facility location problem, routings, etc.Sas; Gupta (2013) .The domination in graphs and signed graphs have been well studied by various authors in different forms viz. Roman domination, double domination, total domination, signed domination, signed total domination etc.Harary and Haynes (2000); Haynes et al. (1998a); Acharya (2013); Ashraf and Germina (2015); Palakkutty et al. (2016); Blidia et al. (2006); Sampathkumar (1991); Zelinka (2001).

In Kansal and Payal, the authors introduced the concept of Signed Petri net(SPN) by utilizing the properties of signed graph and Petri net.SPNs are capable of modeling a large variety of systems and is prefered over a Petri net due to the presence of two types of tokens in it, positive and negative, which are distinguishable. Other advantages of using SPN over previously defined extensions of PN are the ability to assign sign to vertices of an SPN which is further utilized to introduce the concept of a balanced SPN as defined in Kansal and Payal.Further, in comparison to a signed graph, SPN is advantageous since a single SPN can be used to represent various signed graphs by simply varying the marking of SPN due to firing of a sequence of transitions.Thus, we need to analyse one SPN in order to infer about all possible signed graph structures that can be formed for a fixed number of vertices.

As SPN is a bipartite graph, it can be used to develop the theory of domination for dynamic systems as such a theory is not prevalent for Petri nets.

*Email: payal.dtu@gmail.com

†(Corresponding Author)Email: sangita_kansal15@rediffmail.com

ISSN subm. to DMTCS © 2020 by the author(s) Distributed under a Creative Commons Attribution 4.0 International License

2 Basic Definitions

2.1 Petri Net (PN)

A *Petri net* Jensen (1987) is a 5-tuple $N = (P, T, I^-, I^+, \mu_0)$,where

1. P is the finite, non-empty set of places.
2. T is the finite, non-empty set of transitions.
3. $P \cap T = \emptyset$.
4. $I^-, I^+ : (P \times T) \to \mathbb{N}$ where $\mathbb{N}$ is the set of non-negative integers, are called negative and positive incidence functions respectively.
5. $\forall p \in P, \exists\, t \in T$ such that $I^-(p,t) \neq 0$ *or* $I^+(p,t) \neq 0$,and
   $\forall t \in T, \exists\, p \in P$ such that $I^-(p,t) \neq 0$ *or* $I^+(p,t) \neq 0$
6. $\mu_0 : P \to \mathbb{N}$ is the initial marking which gives the initial distribution of tokens in places.

   The arc set of the Petri net N is defined as:

$$E = \{(p,t) : I^-(p,t) > 0\} \cup \{(t,p) : I^+(p,t) > 0\}$$

2.2 Signed Petri Net

Definition 1. ***Signed Petri Net (SPN)***
A Signed Petri Net Kansal and Payal is defined as a 3-tuple $N^* = (N', \sigma, \mu_0)$,where

1. *$N' = (P, T, I^-, I^+)$ is a Petri net structure.*
2. *$\sigma : E \to \{+, -\}$, where E is the arc set of N'.An arc is called a positive or negative arc respectively according to the sign $+$ or $-$ assigned to it using the function σ.*
3. *$\mu_0 = (\mu_0^+, \mu_0^-)$ is the initial marking of SPN where*
   (a) *$\mu_0^+ : P \to \mathbb{N}$ gives the initial distribution of positive tokens in the places,called positive marking of SPN.*
   (b) *$\mu_0^- : P \to \mathbb{N}$ gives the initial distribution of negative tokens in the places,called negative marking of SPN.*

Thus,a marking in SPN can be represented as a vector $\mu = (\mu^+, \mu^-)$ with $\mu^+, \mu^- \in \mathbb{N}^n, n = |P|$ such that $\mu(p_i) = (\mu^+(p_i), \mu^-(p_i)) \; \forall\, p_i \in P$.

Graphically, positive and negative arcs in an SPN are represented by solid and dotted lines respectively.A positive token is represented by a filled circle and a negative token by an open circle.
An SPN is said to be *negative* if all of its arcs are negative in sign.

Remark 1. *$N'' = (N', \sigma)$ is called an SPN structure where N' is a PN structure and $\sigma : E \to \{+, -\}$.*

2.2.1 Execution Rules for Signed Petri Net

Similar to a Petri net, the execution of an SPN depends on the distribution of tokens in its places.The execution takes place by firing of a transition.A transition may fire if it is enabled.
A transition t in an SPN N^* is *enabled* at a marking $\mu = (\mu^+, \mu^-)$ if

$$I^-(p,t) \leq \mu^+(p)\ \forall p \in {}^\bullet t \text{ for which } \sigma(p,t) = +$$

$$I^-(p,t) \leq \mu^-(p)\ \forall p \in {}^\bullet t \text{ for which } \sigma(p,t) = -$$

An enabled transition t may *fire* at $\mu = (\mu^+, \mu^-)$ provided $\exists p_k \in t^\bullet$ such that:

$$\sigma(t,p_k) = \begin{cases} + & \text{if } \sigma(p,t) = + \ \forall p \in {}^\bullet t \\ - & \text{if } \sigma(p,t) = - \ \forall p \in {}^\bullet t \\ + \text{ or } - & \text{if } \sigma(p,t) = + \text{ for some } p \in {}^\bullet t \ and \ - \text{ for some } p \in {}^\bullet t \end{cases}$$

After firing,it yields a new marking $\mu_1 = (\mu_1^+, \mu_1^-)$ given by the rule:

$$\mu_1^+(p) = \mu^+(p) - I^-(p,t) + I^+(p,t)\ \forall p \in P \quad \text{where (p,t) \& (t,p) are positive arcs,}$$

$$\text{if exist}$$

$$\mu_1^-(p) = \mu^-(p) - I^-(p,t) + I^+(p,t)\ \forall p \in P \quad \text{where (p,t) \& (t,p) are negative arcs,}$$

$$\text{if exist}$$

We say that μ_1 is reachable from μ and write $\mu \xrightarrow{t} \mu_1$.We restrict the movement of positive(negative) tokens to positive(negative) arcs only.

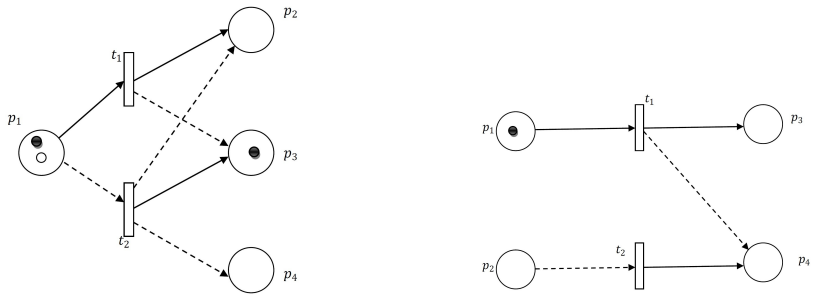
A marking μ is reachable from μ_0 if there exists a firing sequence η that transforms μ_0 to μ and is written $\mu_0 \xrightarrow{\eta} \mu$. A *firing or occurence sequence* is a sequence of transitions $\eta = t_1 t_2 \dots t_k$ such that

$$\mu_0 \xrightarrow{t_1} \mu_1 \xrightarrow{t_2} \mu_2 \xrightarrow{t_3} \mu_3 \dots \xrightarrow{t_k} \mu$$

Note that a transition $t_j, 1 \leq j \leq k$ can occur more than once in the firing sequence η.
Let us look at the execution of an SPN with the help of an example.
In figure $1(a)$, t_1 and t_2 both are enabled at μ_0. Firing of t_1 yields a new marking $\mu = ((0,1,1,0),(1,0,1,0))$ and firing of t_2 yields a new marking $\mu = ((1,0,2,0),(0,1,0,1))$. In figure $1(b)$, t_1 is enabled,while t_2 is not. t_1 can fire to give a new marking $\mu = ((0,0,1,0),(0,0,0,1))$.

Definition 2. ***Reachability Set of Signed Petri net***

The Reachability Set $R(N^*, \mu_0)$ of an SPN N^* is the set of all markings of N^* reachable from the initial marking μ_0.



(a) SPN with $\mu_0 = ((1,0,1,0),(1,0,0,0))$ (b) SPN with $\mu_0 = ((1,0,0,0),(0,0,0,0))$

Fig. 1: Execution of an SPN

2.3 Assignment of sign to vertices of an SPN

The vertices in an SPN can also be assigned sign.Transitions are assigned sign by product of sign of arcs (incoming and outgoing) incident on it.In Figure 1(a) and 1(b), all transitions are negative in sign.

Places can be assigned sign in one of the two ways:

1. **With respect to arcs –** Sign is assigned to a place by taking product of incident arcs (incoming and outgoing) on that place.In Figure 1(a),all the places are negative in sign while in Figure 1(b) p_1 and p_3 are positive in sign while p_2 and p_4 are negative in sign.

2. **With respect to marking –** Sign is assigned to a place by taking product of sign of tokens in that place in the given marking.A place without token is considered to be positive.In Figure 1(a), places p_2, p_3, p_4 are positive in sign while p_1 is negatively signed w.r.t. the marking μ_0.In Figure 1(b), all the places are positively signed w.r.t. the marking μ_0.

Remark– Assigning sign to places with respect to arcs doesn't utilize the most important characteristic(dynamic behaviour) of PN which is a marking.Hence, assigning sign to places with respect to marking has been used throughout the paper.

An example is given which uses the concept of place sign to determine whether a product is approved or disapproved by a company.A company has to make a decision on a certain product by voting of two board members.This situation is represented in figure 2 by modeling it with an SPN.

In the figure, p and q represent the board members .The transitions t_1 and t_3 will fire if positive tokens exist in places p and q while t_2 and t_4 transitions fire if p and q have negative tokens.A positive token is generated in places representing board members(p and q) if they approve the product and else if they dispprove of it , a negative token is generated.A decision on the product is reached if either both p and q disapproves or both approves the product.On the other hand,if one member approves the product while other rejects it,no decision is made.We can determine the decision made by the company on the basis of the sign of the place r in the final marking(when place r gets a token).This is shown in table 1.

Thus,based on the sign of the place r,we can infer the following:

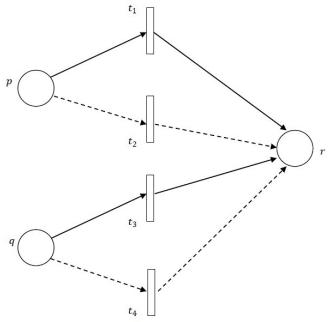


Fig. 2: *An SPN model to check for approval or disapproval of the product*

Tab. 1: Decision Table for the Product

Transition firings	Sign of place r	Decision made
t_1, t_3	+	Yes
t_1, t_4	-	No
t_2, t_3	-	No
t_2, t_4	+	Yes

1. The decision made about the product(The company reaches a final decison if the sign of place r is positive in marking with $\mu(r) \neq 0$,else no decision is made regarding product).

2. Whether p and q have same opinions or varied ones(Both the members have same opinion regarding product if the sign of place r is positive,else they have varied opinion).

If initially the places p and q in figure 2 have one positive and negative token respectively, then this situation corresponds to second row in table 1.Hence,the product will not be approved by the company.

3 Domination Theory

In this section, we assume an ordinary SPN (without multiple arcs) unless stated otherwise.

Definition 3. ***Dominating Set***
A set $D \subseteq V = PUT$ in an SPN N^* is called a Dominating set with respect to a marking $\mu_1 \in R(N^*, \mu_0)$ if either all the vertices of V are in D or $\forall\ v \in V \backslash D$

$$v^\bullet \cap D \neq \emptyset \text{ and } \sigma(v,u) = S(v).S(u)\ \forall\ u \in v^\bullet \cap D$$

where $S(v), S(u)$ are sign of vertices u, v with respect to marking μ_1.

Remark 2. *It may be noted that sign of a transition remains same irrespective of marking of the SPN while sign of a place may vary if the marking of SPN changes.*

Definition 4. ***Dominating Set with respect to a set of markings***(D_M)

A set $D_M \subseteq V$ in an SPN N^* is called a Dominating set with respect to a set of markings $M \subseteq R(N^*, \mu_0)$ if D_M is a dominating set with respect to all the markings $\mu \in M$.(Clearly,$|M| \geq 2$)

Definition 5. ***Dependent(or connected) Dominating Set***

A dominating set D_M with respect to a set of markings M in an SPN N^* is called a dependent Dominating set if the markings of M are all the nodes of some subtree of the reachability tree of N^*.

Definition 6. ***Minimal Dominating Set***

A dominating set D is called Minimal if no proper subset of it is a dominating set or it is a dominating set with minimum number of vertices.

Remark 3. *For application purposes we would like to find minimal dependent dominating set D over a maximal set of markings M,i.e. We try to find a maximal set of markings,M over which D is a minimal dominating set.*

Theorem 1. *For an SPN structure N'' in which all the transitions are positively signed and each place has incident(input/output) arcs of one kind only, we can find a marking μ w.r.t. which $V\backslash A$ is a dominating set where A is a set of source vertices.*

Proof: If $A = \emptyset$, then $V\backslash A = V$ is a dominating set by definition.
If $A \neq \emptyset$, then we have to find a marking μ such that $V\backslash A$ is a dominating set w.r.t. μ. For any $x \in V\backslash(V\backslash A) = A$, $x^\bullet \cap (V\backslash A) \neq \emptyset$(Since, x is a source vertex). Then, we find a marking μ such that $\forall y \in x^\bullet \cap (V\backslash A)$

$$\sigma(x, y) = S(x)S(y) \tag{1}$$

where $S(x)$ and $S(y)$ are sign of vertices w.r.t. μ.

Let $y \in x^\bullet \cap (V\backslash A)$. Now, two cases arise:

1. x **is a place.**
   Therefore, y is a transition. $\implies S(y) = +$ (By hypothesis).
   (a) Now, if $\sigma(x, y) = +$.In this case, for equation 1 to hold ;$S(x) = +$.Therefore, $\mu(x) = (\mu^+(x), \mu^-(x))$ where $\mu^+(x) \in \mathbb{N} \cup \{0\}$ & $\mu^-(x) \in 2\mathbb{N} \cup \{0\}$.
   (b) However, if $\sigma(x, y) = -$.In this case, for equation 1 to hold ;$S(x) = -$.Therefore, $\mu(x) = (\mu^+(x), \mu^-(x))$ where $\mu^+(x) \in \mathbb{N} \cup \{0\}$ & $\mu^-(x) \in \mathbb{N}\backslash 2\mathbb{N}$.
2. x **is a transition.**
   Therefore, y is a place.Then, $S(x) = +$ (By hypothesis).Now, $\sigma(x, y)$ can be positive or negative.
   (a) If $\sigma(x, y) = +$ then, for equation 1 to hold ;$S(y) = +$ and hence, $\mu(y) = (\mu^+(y), \mu^-(y))$ where $\mu^+(y) \in \mathbb{N} \cup \{0\}$ & $\mu^-(y) \in 2\mathbb{N} \cup \{0\}$.
   (b) However, if $\sigma(x, y) = -$ then, for equation 1 to hold ;$S(y) = -$ and hence, $\mu(y) = (\mu^+(y), \mu^-(y))$ where $\mu^+(y) \in \mathbb{N} \cup \{0\}$ & $\mu^-(y) \in \mathbb{N}\backslash 2\mathbb{N}$.

Hence, for all $p_i \in A \cup \{z \in V\backslash A \mid z \in A^\bullet\}$, $\mu^+(p_i) \in \mathbb{N} \cup \{0\}$ &
$\mu^-(p_i) \in \begin{cases} 2\mathbb{N} \cup \{0\} & \text{, if } p_i \text{ has positive incident arcs.} \\ \mathbb{N}\backslash 2\mathbb{N} & \text{, if } p_i \text{ has negative incident arcs.} \end{cases}$
All the remaining places can have any number of positive and negative tokens without any restrictions. □

Theorem 2. *If an SPN structure N'' with no source/sink vertices and in which any place has only one type of incident arcs then, we can find a marking μ such that P and T are dominating sets w.r.t. μ, provided all the transitions are of same sign.*

Proof: Since all the transitions are of same sign,two cases arise:

1. **Transitions are positively signed.**
   We find a marking μ w.r.t. which P and T are dominating sets.

   (a) P **is a dominating set**.
   Let $t \in V \backslash P$,therefore, $t^\bullet \cap P \neq \emptyset$ (since there are no source/sink vertices). We need to find a marking μ such that $\forall\, p \in t^\bullet \cap P$;

   $$\sigma(t,p) = S(t)S(p) \tag{2}$$

   where $S(p)$ is the sign of place p w.r.t. marking μ and $S(t)$ is the sign of transition t.
   Let $p \in t^\bullet \cap P$. Then,

   i. $\sigma(t,p) = +$
   Now, since $S(t) = +\ \forall t \in T$.Then, for equation 2 to hold $S(p) = +$. Then, $\mu_(p) = (\mu^+(p), \mu^-(p))$ where $\mu^+(p) \in \mathbb{N} \cup \{0\}, \mu^-(p) \in 2\mathbb{N} \cup \{0\}$.

   ii. $\sigma(t,p) = -$
   Now, since $S(t) = +\ \forall t \in T$.Then, for equation 2 to hold $S(p) = -$. Then, $\mu_(p) = (\mu^+(p), \mu^-(p))$ where $\mu^+(p) \in \mathbb{N} \cup \{0\}, \mu^-(p) \in \mathbb{N} \backslash 2\mathbb{N}$.

   (b) T **is a dominating set**.
   Let $p \in V \backslash T$,therefore, $p^\bullet \cap T \neq \emptyset$ (since there are no source/sink vertices). We need to find a marking μ such that $\forall\, t \in p^\bullet \cap T$;

   $$\sigma(p,t) = S(p)S(t) \tag{3}$$

   where $S(p)$ is sign of place p w.r.t. marking μ and $S(t)$ is the sign of transition t.
   Let $t \in p^\bullet \cap T$.Then,

   i. $\sigma(p,t) = +$
   Now, since $S(t) = +$. Then, for equation 3 to hold $S(p) = +$. Then, $\mu_(p) = (\mu^+(p), \mu^-(p))$ where $\mu^+(p) \in \mathbb{N} \cup \{0\}, \mu^-(p) \in 2\mathbb{N} \cup \{0\}$.

   ii. $\sigma(p,t) = -$
   Now, since $S(t) = +$. Then, for equation 3 to hold $S(p) = -$. Then, $\mu_(p) = (\mu^+(p), \mu^-(p))$ where $\mu^+(p) \in \mathbb{N} \cup \{0\}, \mu^-(p) \in \mathbb{N} \backslash 2\mathbb{N}$.

   Hence, for all p_i, $\mu^+(p_i) \in \mathbb{N} \cup \{0\}$ &
   $\mu^-(p_i) \in \begin{cases} 2\mathbb{N} \cup \{0\} & , \text{ if } p_i \text{ has positive incident arcs.} \\ \mathbb{N} \backslash 2\mathbb{N} & , \text{ if } p_i \text{ has negative incident arcs.} \end{cases}$

2. **Transitions are negatively signed.**
   By following the same procedure as in the case when transitions are positively signed, we find that

for all p_i, $\mu^+(p_i) \in \mathbb{N} \cup \{0\}$ &

$$\mu^-(p_i) \in \begin{cases} \mathbb{N}\backslash 2\mathbb{N} & , \text{ if } p_i \text{ has positive incident arcs.} \\ 2\mathbb{N} \cup \{0\} & , \text{ if } p_i \text{ has negative incident arcs.} \end{cases}$$

Thus, we get a marking μ w.r.t. which P and T are dominating sets.

□

Remark 4. *The above theorem may be used in case of a disaster to check which transitions representing events in the system will dominate and hence may be used for preparedness against any disaster.*

Remark 5. *Theorems 1 and 2 given above show that:*

1. *We can get a dominating set if we begin with the specified initial marking as mentioned in the proof.*
2. *P & T sets in a PN represent conditions and events respectively of the system modeled.So,we can check whether conditions or events dominate in the given system w.r.t. a given marking.*
3. *If we know the structure of an SPN and marking μ_0, then we can check whether domination can occur by checking if $\exists\ \mu \in R(N^*, \mu_0)$ w.r.t. which there exists a dominating set.So, in order to avoid or force domination we can restrict (or force) SPN to avoid (or reach) such a marking.*

4 Applications of Domination

We discuss applications of domination in SPN that can be utilized in various areas.

4.1 Producer -Consumer Problem

Consider, a standard Producer -Consumer Problem with two producers producing a same product (assuming quality,price and other conditions are same),we need to check whether one of the producer can dominate the market over the other.This can happen due to availability of product is greater for one producer as compared to other or because one product is well known due to its better marketing, etc. Consider an SPN model for the problem given in figure 3.

Now,in order to check if producer 2 dominates the market ,we need to find a set of vertices D_1 which is a dependent dominating set over a maximal set of markings M. A dependent dominating set is considered because one producer is said to dominate over the other if such a domination exists over a period of time, not for just an instant.

Choose $D_1 = V\backslash\{p_7, t_7\}$.Find if there exists a set of markings M(maximal) with marking $\mu \in M$ such that $\mu(p_6) \neq 0$ and D_1 must be a dominating set w.r.t. M. Similarly, to check if producer 1 dominates the market we need to check domination of set D_2 w.r.t. a set of markings where $D_2 = V\backslash\{p_8, t_7\}$.

4.2 Search of food by bees

Bees (Scout bees) go out in search of food.The one which find the food will return to the hive and celebrate.This scout bee can be considered to dominate other scout bees.This problem can be modeled using an SPN and then it can be identified which scout bee will dominate the bee-hive.Consider the figure 4 where place p_1 represents the bee-hive while p_2, p_3 represent possible food locations where scout bees A and B respectively search for food.The transitions t_2, t_3 represent events of food search while t_1, t_4

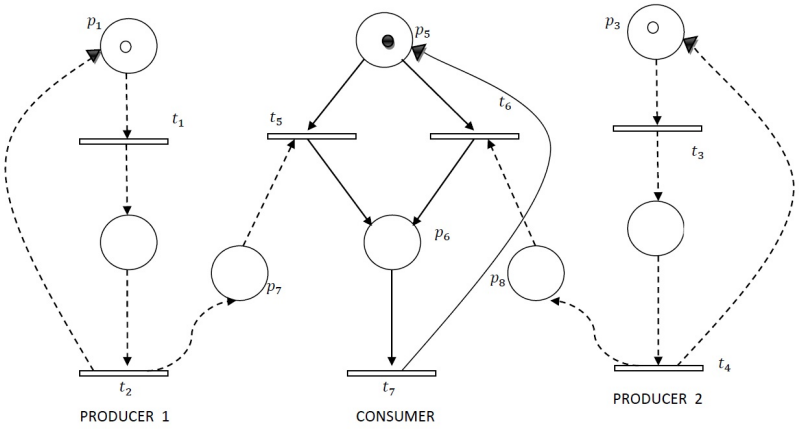


Fig. 3: *Producer-Consumer Problem*

represent events of food search completion.The positive tokens are used to represent bees and negative ones to represent food.

According to the initial marking of SPN, location p_2 has food while location p_3 does not.Therefore, bee A must dominate.This can be verified by checking that the set $D_1 = \{p_1, p_3, t_1, t_2, t_3, t_4\}$ is a dominating set w.r.t.the initial marking μ_0 rather than set $D_2 = \{p_1, p_2, t_1, t_2, t_3, t_4\}$.In the later case, bee B will dominate the bee-hive. The model can be extended if more than two bees search for food.

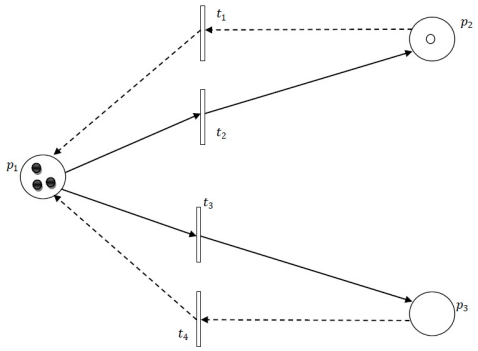


Fig. 4: *An SPN model with initial marking* $((3,0,0),(0,1,0))$ *for food finding by Scout Bees*

4.3 Finding papers with similarity to a given paper using softwares like Turnitin

Consider a paper which need to be checked for similarity using a software.This paper is compared with web content that is publicly available, books, papers in journals, articles and other content which is present

in a software database.Let the paper to be checked be represented by a place p_0 and the rest of the content be represented by places $p_1, p_2, p_3, \ldots p_k$.To check for similarity, an SPN model is formed by connecting place p_0 to all other places via transitions $t_1, t_2, t_3, \ldots t_k$ and by using negative arcs as in figure 5.

While comparing paper p_0 with another article represented by place p_i (say),$1 \leq i \leq k$, a matching algorithm is used to find a set of strings within submitted paper p_0 that matches with the papers maintained in its database.If a similarity exists, a negative token is generated in place p_0 which can be used to fire corresponding transition t_i.In this way, the submitted paper is checked for similarity with all the content present in the database.After the comparisons are completed, we get a new marking for the SPN in which all the articles that have some similarity with the submitted paper get a negative token in the place representing it.All such places will form a list of articles that are similar to the paper submitted which is to be tested for similarity.

Now, in order to find the list of all articles which have some similarity with the submitted paper,the concept of domination can be used instead of finding all the places having a negative token. Begin with set $D_1 = P = \{p_0, p_1, p_2, \ldots p_k\}$.Check whether set D_1 is a dominating set w.r.t. the final marking(say μ') obtained after the matching algorithm is complete.If yes,then all the papers $p_i, 1 \leq i \leq k$ have some similarity with the submitted paper.If not, we find $D_2 \subseteq T$ such that $D' = D_1 \cup D_2$ is a dominating set w.r.t.μ'.Then ,the set $\{p_i | t_i \in T \backslash D_2\}$ will form the set of all the articles which have similarity with the submitted paper.

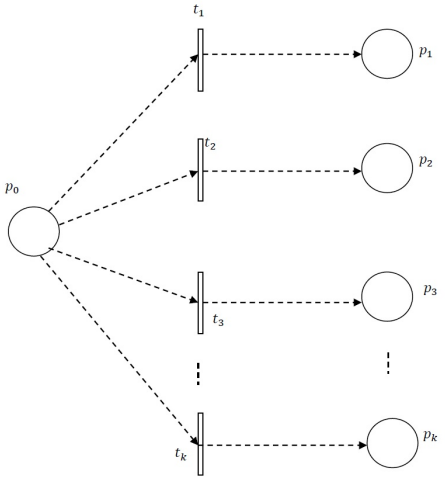


Fig. 5: *Finding similarity in Papers*

References

Applications of dominating set of graph in computer networks. *International Journal of Engineering Sciences and Research Technology*, page 170– 173 year =.

B. D. Acharya. Domination and absorbance in signed graphs and digraphs, i: Foundations. *Journal of Combinatorial Mathematics and Combinatorial Computing*, pages 5–20, 2013.

P. K. Ashraf and K. A. Germina. On minimal dominating sets for signed graphs. *Advances and Applications in Discrete Mathematics*, 15:101–112, 2015.

M. Blidia, M. Chellali, T. W. Haynes, and M. Henning. Independent and double domination in trees. *Utilitas Mathematica*, 70:159–173, 2006.

P. Gupta. Domination in graph with application. *Indian Journal of Research,Paripex*, 2(3):115 – 117, 2013. ISSN 2250-1991.

F. Harary and T. Haynes. Double domination in graphs. *Ars Combinatoria*, 55:201–213, 2000.

T. Haynes, S. T. Hedetneimi, and P. J. Slater. *Fundamentals of domination in graphs*. New York, NY: Marcel Dekker, 1998a.

C. d. Jaenisch. *Trait des Applications de l'Analyse Mathematique au Jeu des Echecs. Petrograd*. 1862.

K. Jensen. Coloured petri nets. *LNCS,Springer, Heidelberg*, 254:248–299, 1987.

S. Kansal and Payal. An introduction to signed petri net. *Communicated*.

A. Palakkutty, A. Germina, and L. Lishan. Double domination in signed graphs. *Cogent Mathematics*, 3:1186135, 2016.

E. Sampathkumar. On some new domination parameters of a graph – a survey. *In Proceedings of a Symposium on Graph Theory and Combinatorics, Kochi, Kerala,India*, pages 17–19, 1991.

B. Zelinka. Signed total domination number of a graph. *Czechoslovak Mathematical Journal*, 51(2): 225–229, 2001.

Environmental Science and Pollution Research
https://doi.org/10.1007/s11356-019-07359-1

ENVIRONMENTAL AND ENERGY MANAGEMENT

Effect of electrostatic precipitator on exhaust emissions in biodiesel fuelled CI engine

Ankit Sonthalia[1,2] · Shivam Garg[3] · Rishav Sharma[1] · Thiyagarajan Subramanian[4] · Naveen Kumar[2]

Received: 21 September 2019 / Accepted: 9 December 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

The exhaust emissions from the compression ignition engines are harmful to both human beings and the environment. After-treatment devices placed in the exhaust are designed to reduce these emissions. These devices have significant conversion efficiency but have various drawbacks such as the cost and availability of the precious catalyst for catalytic converters. In this work, an emission reduction setup was developed that can reduce NO, HC, CO and smoke simultaneously. The emission reduction setup is based on the concept of an electrostatic precipitator (ESP) and plasma generation by corona discharge technique. Both diesel and waste cooking oil biodiesel (WCO) were separately used for the test. The results show that HC emissions at full load with ESP system reduced from 0.71 to 0.27 g/kWh for diesel and for WCO it reduced from 0.81 to 0.31 g/kWh. Similarly, the CO emissions reduced from 1.50 to 0.6 g/kWh for diesel and from 1.95 to 0.92 g/kWh for WCO. The smoke emission and NO emission were also reduced by 30.86 and 29.3% for diesel and WCO and 17 and 18% for diesel and WCO, respectively. However, the carbon dioxide emissions were found to increase as the HC and CO generated were also converted to CO_2. The study shows that the emission reduction setup can effectively reduce the emissions without any effect on the engine performance.

Keywords Electrostatic precipitator · Corona discharge · Plasma · Ozone generation · Emission reduction · Biodiesel

Introduction

The transportation sector and the agricultural sector are dependent upon the diesel engines due to their higher fuel economy, initial torque and low maintenance cost (Geo et al. 2017). However, the main problem with the use of diesel engines is the release of hazardous pollutants in the exhaust namely oxides of nitrogen emissions (NOx), unburnt hydrocarbon emissions (HC), carbon monoxide emissions (CO) and particulate matter. These pollutants not only affect the environment, they also affect the health of human beings (Subramanian et al. 2017). The diesel engines emit particulate matter in the small particle form at a concentration above 10 million particulates per cm^3. The diameter of the primary carbon particles lies in the range of 0.01 to 0.08 μm, whereas the diameter of the agglomerated particles is in the range of 0.08 to 0.25 μm with nearly 90% of the particles having a diameter less than 1 μm. Several government agencies have classified the particulate matter as "human carcinogen" or "probable human carcinogen". They increase the risk to the heart and respiratory diseases such as asthma and bronchitis. Presence of two or more benzene ring hydrocarbons in the particulate matter also called poly-nuclear aromatic hydrocarbons (PAHs) are known human carcinogens. The soluble organic fraction of the particulate matter containing four and five rings is the most harmful compounds (Losacco and Perillo 2018). To tackle the problem of pollutants, the emissions norms all over the world are becoming increasingly stringent. India has also taken a big leap by migrating from Bharat Stage IV (BS-IV) to BS-VI by

Responsible Editor: Philippe Garrigues

✉ Ankit Sonthalia
ankit_sont@yahoo.co.in

[1] Department of Automobile Engineering, SRM Institute of Science and Technology, NCR Campus, Modinagar 201204, India

[2] Centre for Advanced Studies and Research in Automotive Engineering, Delhi Technological University, Bawana Road, Delhi 110042, India

[3] FEV India Private Limited, Pune, India

[4] Green Vehicle Technology Research Centre, Department of Automobile Engineering, SRM Institute of Science and Technology, Kattankulathur 603203, India

Published online: 28 December 2019

2020. The Indian automobile industry now has a new challenge in finding the optimal solution for migrating to BS-VI norms for its existing portfolio of BS-IV norms engine.

The existing exhaust after-treatment devices can effectively reduce the regulated emissions, but there are several drawbacks to their use. The diesel particulate filter placed in the exhaust pipe is used to trap PM emissions; however, the back pressure increases due to its use. There are also problems of regeneration and their efficiency of collecting small particulates is low and the maintenance cost is high (Nguyen Huu Nhon et al. 2004; Stratakis et al. 2002). Similarly, NOx is reduced by selective catalytic reduction (SCR) which has issues like catalyst poisoning, slip of ammonia and low collection efficiency at low loads. These systems are not cost effective and the catalyst used in these systems is not readily available in the aftermarket and if available their cost is prohibitive (Subramaniam et al. 2016).

The electrostatic precipitators (ESP) are used all over the world to control the fly ash in coal-fired power plants. In 1800s, Oliver Lodge gave the idea of ESP and developed it with Alfred Walker for a lead smelter. However, due to high voltage power supply and high resistivity of the lead oxide fumes, the ESP did not work. Later on, in the 1900s, Dr. F. G. Cottrell developed an ESP on industrial scale and patented it (Huang et al. 2015). In ESP a high voltage electric field is generated between two oppositely charged electrodes. The ESP works on the principle of plasma reactions by corona discharge, where electrodes produce high energy electrons which when collides with the molecules present in the air or exhaust gas leads to the formation of new species including ions, atoms and free radicals. The ionized particles migrate towards the oppositely charged electrode. The most significant advantages of ESP is it does not obstruct the flow of air thereby the back pressure is unaffected, since the ionized particles are pulled through the system because of the electrostatic force developed (Parker 1997). Moreover, in the mechanical filters, the gas is passed through some porous media and the gas streamlines are bent based on the separation concept, resulting in pressure drop as soon as the gas hits the surface of the filter (Sudrajad and Yusof 2015). Another advantage is the requirement of less power for operating the ESP (Thonglek et al. 2011). They are widely employed in the industries to collect the fly ash particles.

Many researchers have worked upon removal of soot particles from the diesel engine exhaust using ESP, but no such product is available commercially due to some inherent drawbacks. The particulate matter has low resistance resulting in the particle re-entrainment and hence poor collection efficiency. A two-stage ESP was used by Zukeran et al. (2000). In the first stage, DC was used for charging the particles, and in the second stage AC field with trapezoidal waveforms was used for the collection of particles. Another way of increasing the collection efficiency is the use of agglomerators. Nakajima and Sato (2003) used acoustic waves for strong vibration of the particles along with electrostatic polarization of the particles. Due to acoustic waves, the smaller particles move relatively higher than coarser particles which increase the chance of collision, thereby increasing the agglomeration. Ji et al. (2004) agglomerated the particles by first bipolar charging and then passing the particles through an AC electric field and a separate DC electric field. The authors observed 25–29% reduction in particles smaller than 1 μm with AC frequency. Ozone regeneration by corona discharge is one of the best methods to incinerate the particles deposited on the electrode (Kuwahara et al. 2013; Umemoto et al. 2008). Subramaniam et al. (2016) designed an ESP containing spike-shaped charging electrode and a three-layered collecting electrode. The authors used a corona discharge technique for the generation of ozone resulting in the simultaneous reduction of HC, NOx and smoke.

Many methods have been used for removal of NOx such as injection of a radical for NOx oxidation and reduction. Some of the radicals are ammonia (Nishida et al. 2001), ozone (Okubo et al. 2008; Sano and Yoshioka 2003), methane (Chang et al. 1998; Kanazawa et al. 1997), etc. Another method is the use of plasma for reducing NOx in a gas stream. In one study, a system consisting of plasma and Ag, Au/Al_2O_3 catalyst was used for reducing NOx (Ryu et al. 2002). With model exhaust gas, the NOx reduction efficiency varied from 40 to 90% in the temperature range of 250 to 400 °C. With an engine test, the NOx removal efficiency was 46% at 364 °C. The system also showed remarkable NOx reduction tendency at 100 to 250 °C. In another study, non-thermal plasma along with SCR has been used to enhance the rate of NOx removal (Sato et al. 2011). The ammonia gas was supplied in the SCR system. With simulated gas of a diesel engine exhaust, large NOx reduction was observed. The authors also tested the system on a heavy duty engine over JE05 Japanese cycle and observed significant NOx reductions. On the laboratory scale, also numerous studies were done to reduce the NOx level (Talebizadeh et al. 2014; Tonkyn et al. 2003; Tran et al. 2004; Ueda et al. 2004). Almost all these studies use plasma along with lean burn internal combustion engines.

The literature shows that the existing after-treatment systems are effective in reducing emissions but they have their own drawbacks. Such as the presence of sulphur in the fuel poisons the catalysts present in the after-treatment system. The diesel particulate filter has various issues like high back pressure, regeneration and poor collection efficiency of small-sized particles. In the present study a plasma-assisted exhaust after-treatment system was developed for a single cylinder diesel engine. The system contains one electrostatic precipitator and one non-thermal plasma generator working on pulse corona discharge technique. The ESP was used for reducing the exhaust smoke and the plasma generator was used to reduce other emissions. The novelty of the work lies in the use

of Ni/Al_2O_3 catalyst which has not been used so far by any of the researchers. The catalyst was coated on the surface of the negative electrode of the plasma generator for improving the emission conversion efficiency. The study aims to simultaneously reduce the hazardous pollutants from the engine exhaust without the drawbacks of the existing after-treatment systems and without affecting the engine performance. The engine tests were conducted with diesel, and to ascertain that the after-treatment system works well with fuels other than diesel, engine tests were also conducted with neat waste cooking oil biodiesel.

Waste cooking oil biodiesel production

Waste cooking oil is readily available all over the world. It can be used as an additive for the preparation of fodder and in the soap manufacturing. However, vast quantities of WCO are either dumped in rivers or wastelands thus harming the environment (Senthil Kumar and Jaikumar 2014). In India, there are many roadside shops which use cooking oil for making food. The shop owners use the oil two to three times before throwing it in the drains which are harmful both for the humans by eating the food prepared in reused oil and the environment as the used oil is thrown in the drain. Instead, the oil can be converted to biodiesel and used as fuel in internal combustion engines (Arslan and Ulusoy 2018).

Waste cooking oil was collected from roadside shop owners in and around Delhi. The oil was initially filtered to remove the impurities and heated to 110 °C so that all the water (if present) is evaporated. The purified WCO was placed in a batch-type reactor which is a round bottom flask having three necks. The central neck has a stirrer connected to a DC motor. One of the necks has a thermometer so that the temperature of the reactants can be continuously monitored. The flask was placed on a heating jacket for heating the flask so that the temperature of the reactants was maintained accordingly. A mixture of 185 g of methanol and 4.6 g of potassium hydroxide (optimized) was prepared separately in a beaker. The mixture was then added to the reactant under constant stirring and 65–70 °C was maintained for 45 min. After the reaction was complete, the mixture was settled for 12–24 h in a separating funnel. After 24 h, two layers were formed: the lower layer was darker in colour and it consisted of glycerine and other products, while the upper layer was light in colour and consisted mainly of biodiesel. The two layers were then separated and the methyl ester of waste cooking oil was purified by washing with water. The biodiesel was heated to above 100 °C, to evaporate any water left. The conversion rate of waste cooking oil into biodiesel was nearly 92%. Some of the properties of the obtained biodiesel are shown in Table 1.

Construction and working of the ESP

The after-treatment setup consists of two electrostatic precipitators and a catalyst layer attached to the negative electrode of the second ESP. A gasket was placed in between the electrodes and the outer pipe. The gasket acts as an insulator. Figure 1 shows the schematic of the ESP. Nickel supported on aluminium oxide was used as the catalyst. The catalyst was made by preparing nickel and aluminium precursor (salt) solutions in ethanol. The nickel-aluminium solution was then coated on thin aluminium plates pre-coated with silica. The plates were then dried at room temperature followed by calcination at 350 °C.

The exhaust gas first passes through the wire mesh with negative potential. The particles become negatively charged by passing through the electric field of the wire mesh. As the positive electrode is placed just after the wire mesh, the particles are attracted to the positive electrode which is cylindrical in shape. The positive electrode acts as a collector wherein some of the negatively charged particles are collected. In the second precipitator, a rod is placed in the centre of the pipe, and the rod has positive potential. The catalyst is placed on the surface of the negative electrode. Positive polarity pulse is used in the present work, since in the earlier works also positive polarity resulted in higher removal efficiency of gaseous pollutants (Kinoshita 2002; Thomas et al. 2001). As the streams take longer time in a positive pulse corona discharge, the ionization of more substantial volume of gas takes place in comparison to negative pulse corona discharge resulting in better performance (Li et al. 2000). A 13-kV AC power was applied between the two electrodes which helped in the generation of ozone (for few seconds). As the particles are attracted towards the negative layer, the catalyst present there provides the surface area for the oxidation of carbon monoxide and particulate matter, and reduction of hydrocarbons and oxides of nitrogen by ozone.

Experimental setup

In the present work, a Kirloskar make TAF-1 engine is used. The engine is four-stroke, air-cooled, direct injection, single cylinder; its specifications are given in Table 2. The engine was loaded in steps of 25% of full load using a DC generator connected to a load bank. To switch between diesel and WCO, a two-position, three-way manually operated directional control valve was used. An AVL di-gas analyser was used to measure NO, CO, CO_2 and HC emissions in the exhaust gas and an AVL smoke meter was used for measuring the smoke opacity (Fig. 2).

The experiments were conducted at a constant engine speed of 1500 rpm and the loads were varied in steps of 25% of 100% load. Using a 50-cm^3 burette and a stopwatch,

Table 1 Fuel properties

Fuel property	ASTM standard	Equipment used	Diesel	WCO biodiesel
Density at 15.5 °C kg/m^3	ASTM D4052	Anton Parr DMA 4500	840	857
Kinematic viscosity at 40 °C cSt	ASTM D445	Capillary tube	2–6	5.7
Flash point °C	ASTM D93	Pensky-Martens (AD0093-600)	69	110
Gross calorific value in MJ/kg	ASTM D240	Parr 6100 Oxygen Bomb Calorimeter	44.8	39.6
Calculated Cetane Index (CCI)	–	–	47	45

the volumetric fuel flow rate was measured. Brake thermal efficiency and exhaust emissions were used to compare the engine performance. In the first phase, baseline was formed by conducting the experiments without the ESP. In the second phase, the experiments were conducted with the ESP. Only when the engine reached steady state condition, the readings were taken in triplicate and averaged. The uncertainty of the instruments used in this study is given in Table 3. The total uncertainty of the experiment was found using propagation of error techniques and was observed to be ± 3.2%. Appendix 1 shows the sample calculation for the uncertainty analysis.

Results and discussion

Figure 3 shows the effect of load and electrostatic precipitator on NO emission. The oxides of nitrogen are formed due to the high temperature of combustion which facilitates oxidation of nitrogen present in the air (Varuvel et al. 2018a). The NO emission level is higher with diesel fuel engine operation as compared to WCO engine operation at all loads. At full load, the NO emission with diesel and WCO engine operation is 7.26 and 7.01 g/kWh, respectively. The NO is formed mainly during the premixed combustion phase. WCO has lower NO emission since the premixed combustion phase is reduced, due to poor spray characteristics resulting in low fuel/air mixture. Hence less fuel is prepared for combustion during the initial phase (Boopathi et al. 2018). When the ESP is placed in the exhaust pipe, at full load, the NO emission decreases by 17% with diesel, and 21% with WCO as compared to engine operation with baseline diesel without ESP. The reduction of NOx occurs in two steps. First, the plasma oxidizes the NO to NO_2 in the presence of hydrocarbons. The reaction for the oxidation is given by Eq. 1 (Penetrante et al. 1998; Puchkarev et al. 1999).

$$\text{plasma} + \text{NO} + \text{HC} + \text{O}_2 \rightarrow \text{NO}_2 + \text{HC}-\text{products} \quad (1)$$

Here HC refers to hydrocarbons and HC-products refers to hydrocarbons that are partially oxidized. In the second step, as the formed NO_2 and products of HC are attracted towards the negative potential, the catalyst with the help of hydrocarbons convert NO_2 to N_2 using Eq. 2.

$$\text{catalyst} + \text{NO}_2 + \text{HC} \rightarrow \text{N}_2 + \text{CO}_2 + \text{H}_2\text{O} \quad (2)$$

Both NO_2 and HC are absorbed on the surface of the catalyst wherein the hydrocarbon species present near to the sites where NO_2 is absorbed chemically removes the oxygen through the formation of transient species and then the nitrogen molecule exits the catalyst. In the process, the hydrocarbon is converted to carbon monoxide and carbon dioxide. The basic chemistry of the reaction is thus confirmed, although there is no clarity on the possible transient species formed. The reaction also shows that the presence of sufficient

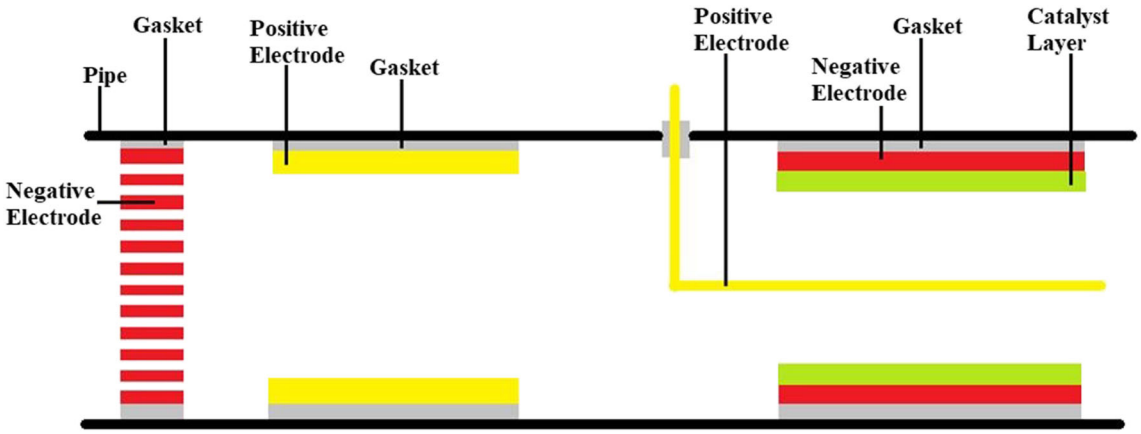


Fig. 1 The electrostatic precipitator

Table 2 Specification of the test engine

Make	Kirloskar
Type	Single cylinder, CI, 4-stroke
Type of cooling	Air cooled
Bore × stroke	87.5 × 110 mm
Compression ratio	17.5:1
Piston bowl	Hemispherical
Rated power	4.4 kW at 1500 rpm
Lubrication oil	SAE 40

hydrocarbon level is necessary for the reactions to occur. Theoretically equal amount of NO and HC are required for the reaction. However, higher ratios of hydrocarbon will increase the probability of NO reduction (Herling et al. 2000). Figure 3 also shows a higher reduction of NO emissions with WCO engine operation as compared to diesel engine operation mainly due to the presence of higher amount of unburnt hydrocarbons in the exhaust gas (Fig. 4) which helps in the reduction of NO by the process previously discussed.

The variation in HC emission with load for the test fuels is portrayed in Fig. 4. The HC emission with WCO engine operation is higher than diesel operation due to incomplete combustion. The density and viscosity of WCO are higher than diesel resulting in improper penetration of the spray thereby causing poor mixing of fuel and air and formation of partially burnt hydrocarbons (Geo et al. 2018). The HC emission level at full load without the electrostatic precipitator for diesel and WCO is 0.71 and 0.82 g/kWh, respectively. With ESP in place, the emission level reduces for both the test fuels. The HC emissions with ESP reduced to 0.27 and 0.31 g/kWh for diesel and WCO engine operation, respectively. The emission reduction with ESP can be attributed to the use of HC emissions for converting NO to N_2, as previously discussed. Also, the emissions are oxidized during the ozone regeneration process in the plasma along with the catalyst, which speeds up the oxidation process, to carbon monoxide according to Eq. 3 (Subramaniam et al. 2016).

$$2HC + 5O \rightarrow 2CO_2 + H_2O \quad (3)$$

The CO emissions are generally lower in CI engines as compared to SI engines as the CI engines run on lean mixtures. Partial oxidation of the fuel molecules and reduced flame temperatures are the primary reasons for the formation of the intermediate carbon monoxide (Thiyagarajan et al. 2019). The variation in CO emission with load for the conducted tests is shown in Fig. 5. At 100% load, the CO emission is 1.5 and 1.95 g/kWh for diesel and WCO, respectively. The fuel molecules unburnt during the initial stages are surrounded by their own products of combustion in the later stages of combustion thus they do not have enough oxygen for their oxidation. Also, by the end of the expansion stroke, the CO that has formed does not have enough time for its oxidation along with lower temperature (Geo et al. 2017). As the ESP is placed in the exhaust pipe, the carbon monoxide emission levels reduced. At full load, as compared to diesel operation without ESP, the emission levels reduced by 60 and 40% with diesel and WCO engine operation, respectively. In the plasma, generated by the corona discharge, the ozone is formed according to the Eqs. 4 and 5. The ozone reacts with the carbon monoxide present in the exhaust gas according to the Eqs. 6 and 7 (Li et al. 2000), thus forming carbon dioxide. Figure 5 confirms that the complete conversion of carbon monoxide is not possible probably due to less residence time.

$$O_2 + e^- \rightarrow 2O + e^- \quad (4)$$

$$O_2 + O \rightarrow O_3 \quad (5)$$

$$CO + O_3 \rightarrow CO + O_2 + O \quad (6)$$

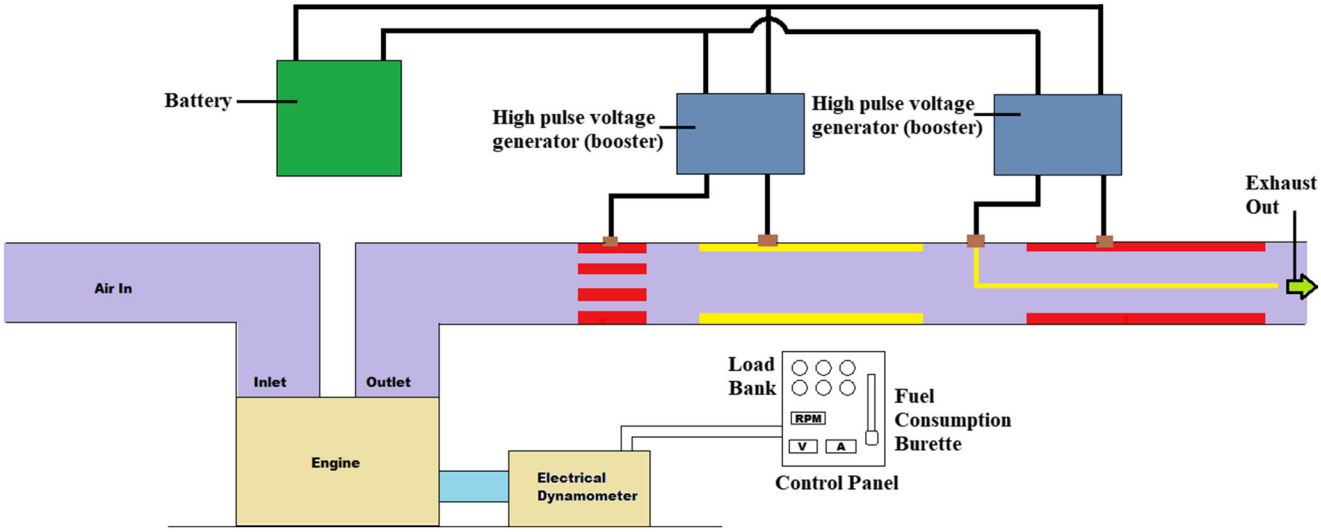


Fig. 2 Schematic of the experimental setup

Table 3 Uncertainty of different instruments

Measurement	Accuracy	% Uncertainty	Measurement technique
Voltage	± 1 V	± 0.2	Voltmeter
Ammeter	± 0.1 A	± 0.3	Ammeter
Speed	± 10 rpm	± 0.1	Magnetic pickup
Burette for fuel measurement	± 0.1 cc	± 1	Volumetric measurement
Time	± 0.1 s	± 0.2	Manual stop watch
CO emission	± 0.02%	± 0.2	NDIR principle
HC emission	± 20 ppm	± 0.2	NDIR principle
CO_2 emission	± 0.03%	± 0.15	NDIR principle
NO	± 10 ppm	± 1	Electrochemical measurement
Smoke	± 1%	± 1	Light extinction method
Exhaust gas temperature	± 1 °C	± 0.15	K-type thermocouple

$$CO + O \rightarrow CO_2 \quad (7)$$

The variation in carbon dioxide emissions with load is shown in Fig. 6. Carbon dioxide is formed when the fuel containing hydrocarbons is completed burned. The amount of carbon that the fuel contains and the fuel consumption play a significant role in the formation of CO_2 (Thiyagarajan et al. 2018a, b). The carbon dioxide emissions are lower with waste cooking oil biodiesel. Although WCO has higher oxygen content and the fuel consumption is also higher, the CO_2 emissions are lower on account of slower combustion due to high viscosity and lower combustion temperature. The CO_2 emissions at full load for diesel and WCO engine operation without ESP are 8.19 and 7.28 g/kWh, respectively. Whereas, with ESP placed in the exhaust pipe, the carbon dioxide emission at full load is 9.42 and 8.86 g/kWh with WCO and diesel engine operation, respectively. The carbon dioxide emission without the emission reduction setup is lower than with the setup in place. The plasma generated in the corona discharge setup and the catalyst placed on the negative electrode of the corona discharge setup facilitate the conversion of unburned hydrocarbons and carbon monoxide to carbon dioxide, according to Eqs. 2, 3 and 7, thereby increasing the CO_2 emission levels.

The modulation in smoke opacity with various engine loads is shown in Fig. 7. The combustion in diesel engines is heterogeneous resulting in air-fuel ratio being rich in some parts of the combustion chamber. The fuel molecules in these regions do not get enough oxygen and are thermally cracked resulting in the formation of large amount of smoke (Edwin Geo et al. 2019). Waste cooking oil biodiesel engine operation resulted in higher smoke opacity than diesel on account of large quantity of fuel injected into the combustion chamber due to its lower calorific value. Also, the viscosity of the fuel is high and its volatility is poor which makes the combustion inferior. Combining these two effects results in the partial oxidation of the carbon atoms thereby increasing the formation of particulate matter. The smoke opacity is lower with the emission reduction system in place. The first electrostatic precipitator tends to reduce the particulate emission by charging

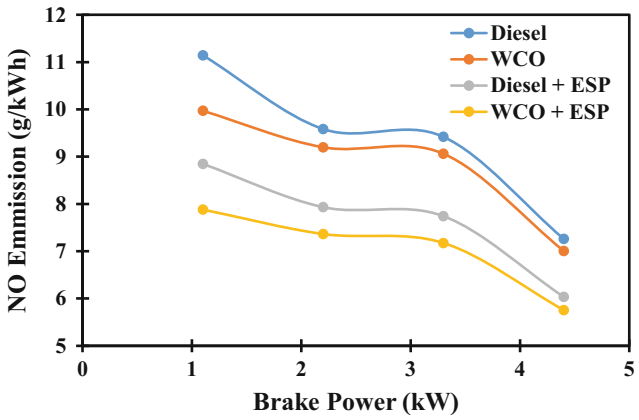


Fig. 3 Variation in oxides of nitrogen emission with engine load

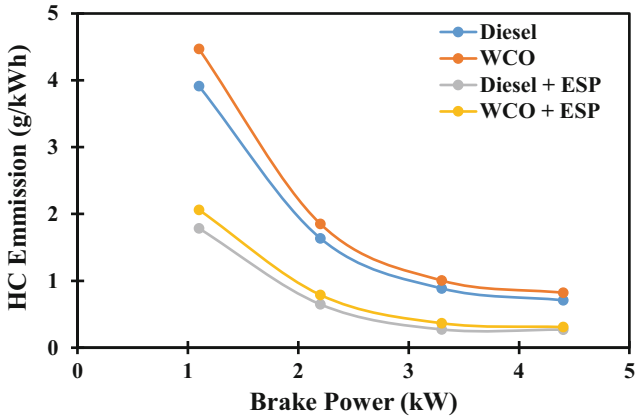


Fig. 4 Variation in unburned hydrocarbon emission with engine load

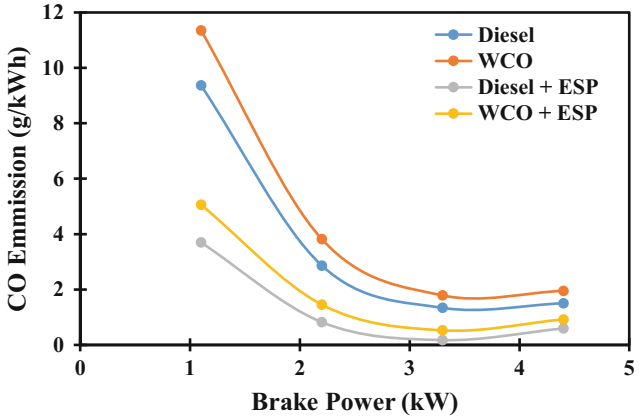


Fig. 5 Variation in carbon monoxide emission with engine load

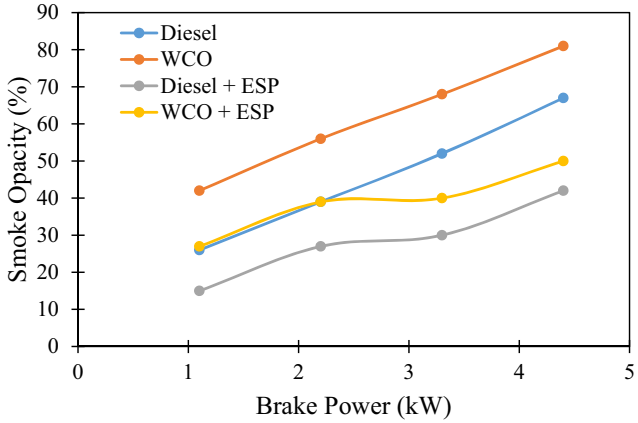


Fig. 7 Variation in opacity of the smoke with engine load

the particles which are collected on the positive electrode of the electrostatic precipitator. When the exhaust gas moves further into the second ESP, the ozone generated in the plasma oxidizes some of the carbon particles according to Eq. 8.

$$C + O_3 \rightarrow CO_2 + \frac{1}{2}O_2 \tag{8}$$

Figure 7 also shows that the particulate reduction efficiency is high at higher loads in comparison to lower loads. It is observed that at 25% load the particulates reduction efficiency is 30.7 and 30.3% for diesel and WCO, respectively. Whereas at full load the reduction efficiency increases to 37.3 and 38.2% for diesel and WCO, respectively. With varying load conditions, a significant change occurs in the concentration of mass and composition of the particles. The particle mass concentration at high loads is less, but the number concentration is high resulting in a shift of size distribution towards smaller sized particles (Subramaniam et al. 2016). The data observed shows that the efficiency of ESP in collecting the small-sized particles is more in comparison to relatively bigger sized particles.

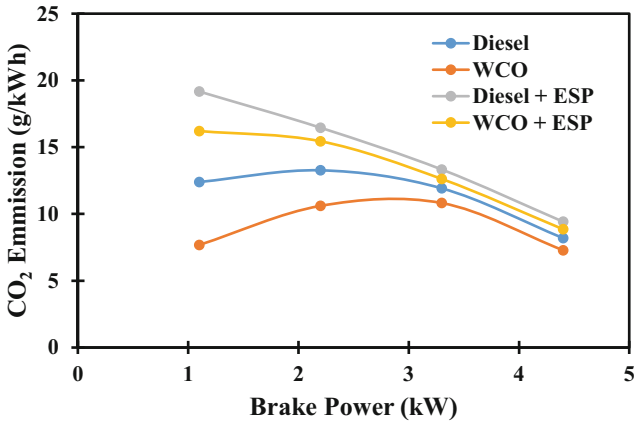


Fig. 6 Variation in carbon dioxide emission with engine load

The engine efficiency is expected to decrease with catalytic converter placed in the exhaust pipe due to the back pressure caused by the obstruction to the flow of the exhaust gas. The advantage that the emission reduction setup used in the present work has over conventional catalytic converter is that the exhaust gas is not obstructed and the engine efficiency is the same, shown in Fig. 8. Moreover, the brake thermal efficiency with waste cooking oil engine operation is lower than that of diesel. At full load, the efficiency is 30.86 and 29.3% for diesel and WCO, respectively. The decrease in efficiency is due to poor spray characteristics caused by high viscosity resulting in poor mixing leading to inferior combustion (Varuvel et al. 2018b). Therefore, the heat released in premixed combustion is less due to longer ignition delay resulting in slow combustion and the flame front reaching the exhaust valves (Thiyagarajan et al. 2018a, b).

The variation in exhaust gas temperature (EGT) with engine load is shown in Fig. 9. It is seen that the EGT increases as the load on the engine increases. WCO engine operation resulted in higher exhaust temperature than diesel engine operation. The increase in EGT with WCO is due to higher heat release in the diffusion combustion phase which is not utilized in useful work but leaves the engine as waste heat (Boopathi et al. 2018). The exhaust temperature was similar with after-treatment system and without the system for both the fuels.

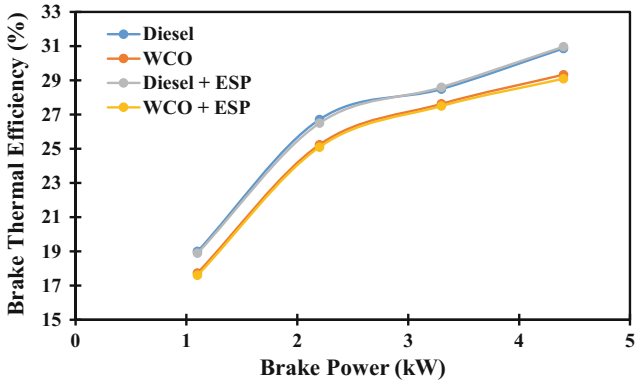


Fig. 8 Variation in brake thermal efficiency with engine load

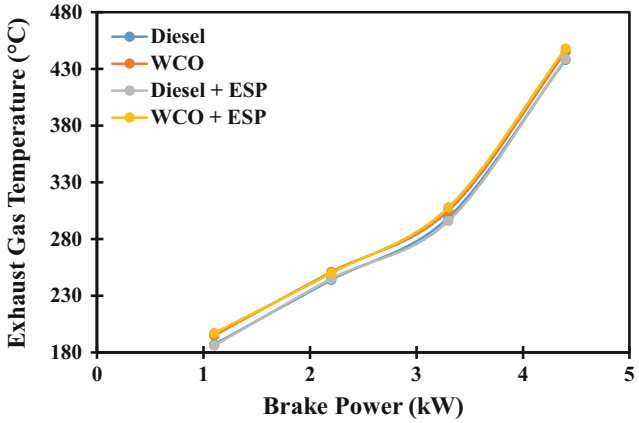


Fig. 9 Variation in exhaust gas temperature with engine load

This shows that the after-treatment system has no effect on the EGT. However, as shown in Fig. 3, the NO removal efficiency of the system decreases as the load on the engine increases. This effect is mainly attributed to the reduction in HC emissions at high loads (and high EGT), since HC emissions play a role in the reduction of NO emission. It is also seen that there is some improvement in the conversion efficiency of other emissions, since the catalyst efficiency improves with the increase in temperature. Therefore, it can be considered that temperature of the exhaust gas plays minuscule role in the working of the after-treatment system.

Conclusions

In the present work, the ESP was developed to reduce the pollutants namely HC, CO, NO and smoke from the exhaust of a compression ignition engine fuelled with diesel and WCO. The results show that NO emission is reduced by 20.8 and 16.9% for WCO and diesel, respectively, with ESP placed in the exhaust line. The plasma formed in the ESP along with unburnt hydrocarbons in the exhaust gas helps in the conversion of NO to nitrogen. The catalyst placed on the negative electrode of the ESP facilitates the faster conversion of NO to N_2. The HC emissions are reduced by 61.9 and 62.5% with diesel and WCO engine operation. The ozone present in the plasma oxidizes the unburnt hydrocarbon to carbon dioxide along with the use of HC emissions in converting NO to N_2, thereby reducing the HC emissions. The ozone and the catalyst also facilitate the faster conversion of CO to CO_2. At full load, the CO emission reduced by 60.4 and 52.89% with diesel and WCO engine operation, respectively. Since the unburnt hydrocarbons and carbon monoxide are converted to CO_2 by the emission reduction setup, the emission level at full load increased by 9.18 and 9.89% with WCO and diesel engine operation, respectively. The particulate matter emission is also reduced by the ESP. The removal efficiency is higher at high loads when the particle size is small. The brake thermal efficiency with and without the ESP shows that the ESP does not affect the engine performance. Thus it can be concluded that the ESP can reduce the pollutants from the exhaust gas of an engine without any modifications of the engine.

Future scope and limitation

In an electrostatic precipitator, due to the deposition of the smoke particles on the electrodes, its efficiency starts decreasing after certain hours of running. Therefore, it is necessary to regenerate the electrodes so that the ESP works with the same efficiency. In the present study, the efficiency of the ESP was not affected as after every test the electrodes were cleaned. Similarly, the efficiency of the catalyst can reduce after certain period of time. The future scope of the study is to devise a mechanism through which the electrodes of the ESP can be regenerated after certain hours of running of the vehicle. Also, the amount of energy required for operating an ESP is generally high. Therefore, a study can be carried out for finding the energy spent on operating the ESP and the optimization of the power required so that a balance exists between the emission reduction efficiency and the energy requirement of the system.

Acknowledgements The authors are thankful to Dr. B. P. Kiran from Synthesis with Catalysts Pvt. Ltd. for providing the catalyst.

Appendix 1

Sample calculation for the uncertainty analysis

Uncertainty in brake power = $\sqrt{\left(\frac{\Delta V}{V}\right)^2 + \left(\frac{\Delta I}{I}\right)^2} = 1.2\%$

Uncertainty in fuel consumption = $\sqrt{\left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta t}{t}\right)^2} = 1.5\%$

Uncertainty in brake thermal efficiency = $\sqrt{\left(\frac{\Delta BP}{BP}\right)^2 + \left(\frac{\Delta FC}{FC}\right)^2}$ = 1.12%

Uncertainty in oxides of nitrogen = $\sqrt{\left(\frac{\Delta BP}{BP}\right)^2 + \left(\frac{\Delta NO}{NO}\right)^2}$ = 1.1%

Uncertainty in carbon monoxide = $\sqrt{\left(\frac{\Delta BP}{BP}\right)^2 + \left(\frac{\Delta CO}{CO}\right)^2}$ = 1.12%

Uncertainty in unburnt hydrocarbon = $\sqrt{\left(\frac{\Delta BP}{BP}\right)^2 + \left(\frac{\Delta HC}{HC}\right)^2}$ = 1.14%

Uncertainty in carbon dioxide = $\sqrt{\left(\frac{\Delta BP}{BP}\right)^2 + \left(\frac{\Delta CO_2}{CO_2}\right)^2}$ = 1.15%

Uncertainty in smoke = 1.0%

Where V—the voltage available from the generator, I—the current available from the generator, A—graduated burette, t—time, BP—brake power, FC—fuel consumption, NO—

oxides of nitrogen, CO—carbon monoxide, HC—unburnt hydrocarbon, and CO_2—carbon dioxide.

References

Arslan R, Ulusoy Y (2018) Utilization of waste cooking oil as an alternative fuel for Turkey. Environ Sci Pollut Res 25(25):24520–24525. https://doi.org/10.1007/s11356-017-8899-3

Boopathi D, Thiyagarajan S, Edwin Geo V, Madhankumar S, Gheith R (2018) Effect of geraniol on performance, emission and combustion characteristics of CI engine fuelled with gutter oil obtained from different sources. Energy 157:391–401. https://doi.org/10.1016/J.ENERGY.2018.05.129

Chang JS, Urashima K, Arquilla M, Ito T (1998) Reduction of NOx from combustion flue gases by corona discharge activated methane radical injections. Combust Sci Technol 133(1–3):31–47. https://doi.org/10.1080/00102209808952025

Edwin Geo V, Sonthalia A, Nagarajan G, Nagalingam B, Aloui F (2019) Effect of port premixed liquefied petroleum gas on the engine characteristics. J Energy Resour Technol 141(11). https://doi.org/10.1115/1.4043698

Geo VE, Sonthalia A, Nagarajan G, Nagalingam B (2017) Studies on performance, combustion and emission of a single cylinder diesel engine fuelled with rubber seed oil and its biodiesel along with ethanol as injected fuel. Fuel 209:733–741. https://doi.org/10.1016/j.fuel.2017.08.036

Geo VE, Sonthalia A, Aloui F, Femilda Josephin JS (2018) Study of engine performance, emission and combustion characteristics fueled with diesel-like fuel produced from waste engine oil and waste plastics. Front Environ Sci Eng 12(4):8–9. https://doi.org/10.1007/s11783-018-1063-6

Herling D, Smith M, Baskaran S, Kupe J (2000) Application of non-thermal plasma assisted catalyst technology for diesel emission reduction. SAE Technical Paper Series 2000-01–3088. https://doi.org/10.4271/2000-01-3088

Huang Y, Huang S, Zheng Q, Shen X, Wang S, Han P et al (2015) Recent progress of dry electrostatic precipitation for PM2.5 emission control from coal-fired boilers. International Journal of Plasma Environmental Science and Technology 9(2):69–126

Ji J-H, Hwang J, Bae G-N, Kim Y-G (2004) Particle charging and agglomeration in DC and AC electric fields. J Electrost 61(1):57–68. https://doi.org/10.1016/j.elstat.2003.12.003

Kanazawa S, Chang JS, Round GF, Sheng G, Ohkubo T, Nomoto Y, Adachi T (1997) Removal of NOx from flue gas by corona discharge activated methane radical showers. J Electrost 40–41:651–656. https://doi.org/10.1016/S0304-3886(97)00072-7

Kinoshita K (2002) Study on removal of PM and NOx in diesel exhaust by using DC corona discharge. SAE Technical Paper 2002-01–16(724). https://doi.org/10.4271/2002-01-1660

Kuwahara T, Nishii S, Kuroki T, Okubo M (2013) Complete regeneration characteristics of diesel particulate filter using ozone injection. Appl Energy 111:652–656. https://doi.org/10.1016/J.APENERGY.2013.05.041

Li X, Yang L, Lei Y, Wang J, Lu Y (2000) A method for removal of CO from exhaust gas using pulsed corona discharge. J Air Waste Manag Assoc 50(10):1734–1738. https://doi.org/10.1080/10473289.2000.10464215

Losacco C, Perillo A (2018) Particulate matter air pollution and respiratory impact on humans and animals. Environ Sci Pollut Res 25(34):33901–33910. https://doi.org/10.1007/s11356-018-3344-9

Nakajima Y, Sato T (2003) Electrostatic collection of submicron particles with the aid of electrostatic agglomeration promoted by particle vibration. Powder Technol 135–136:266–284. https://doi.org/10.1016/S0032-5910(03)00165-7

Nguyen Huu Nhon Y, Mohamed Magan H, Petit C (2004) Catalytic diesel particulate filter: evaluation of parameters for laboratory studies. Appl Catal B Environ 49(2):127–133. https://doi.org/10.1016/J.APCATB.2003.12.005

Nishida M, Yukimura K, Kambara S, Maruyama T (2001) Reduction of nitrogen oxide in N2 by NH3 using intermittent dielectric barrier discharge. J Appl Phys 90(6):2672–2677. https://doi.org/10.1063/1.1394902

Okubo M, Arita N, Kuroki T, Yoshida K, Yamamoto T (2008) Total diesel emission control technology using ozone injection and plasma desorption. Plasma Chem Plasma Process 28(2):173–187. https://doi.org/10.1007/s11090-008-9121-7

Parker KR (1997) Milestones in the history of precipitation. In: Parker KR (ed) Applied electrostatic precipitation. Blackie Acad.. & Professional, London, pp 11–12

Penetrante BM, Brusasco RM, Merritt BT, Pitz WJ, Vogtlin GE (1998) Plasma-assisted catalytic reduction of NOx. SAE Technical Paper Series 982508

Puchkarev V, Kharlov A, Gundersen M, Roth G (1999) Application of pulsed corona discharge to diesel exhaust remediation. Digest of Technical Papers. 12th IEEE International Pulsed Power Conference. (Cat. No.99CH36358) 1(October):511–514. https://doi.org/10.1109/PPC.1999.825522

Ryu J, Jeong D, Chun B, Lee D, Lee K, Lee H, Chun KM (2002) Experimental study on DeNOx performance by plasma-catalyst (Ag, Au/Al2O3) system. SAE International (2002-01–2705)

Sano K, Yoshioka Y (2003) Effect of oxygen injection and HC addition on NO removability of ozone injection method. IEEJ Trans Fundam Mater 123(10):1030–1036. https://doi.org/10.1541/ieejfms.123.1030

Sato S, Kawada Y, Sato S, Hosoya M (2011) The study of NOx reduction using plasma-assisted SCR system for a heavy duty diesel engine. SAE International (2011-01–0310). https://doi.org/10.4271/2011-01-0310

Senthil Kumar M, Jaikumar M (2014) A comprehensive study on performance, emission and combustion behavior of a compression ignition engine fuelled with WCO (waste cooking oil) emulsion as fuel. J Energy Inst 87(3):263–271. https://doi.org/10.1016/j.joei.2014.03.001

Stratakis GA, Psarianos DL, Stamatelos AM (2002) Experimental investigation of the pressure drop in porous ceramic diesel particulate filters. Journal of Automobile Engineering 216:773–784

Subramaniam MK, Pachamuthu S, Arulanandan J, Muthiya J (2016) Simultaneous reduction of HC, NOx and PM by using active regeneration technique. SAE International 2016-01–0912

Subramanian T, Varuvel EG, Martin LJ, Beddhannan N (2017) Effect of lower and higher alcohol fuel synergies in biofuel blends and exhaust treatment system on emissions from CI engine. Environ Sci Pollut Res 24(32):25103–25113. https://doi.org/10.1007/s11356-017-0214-9

Sudrajad A, Yusof AF (2015) Review of electrostatic precipitator device for reduce of diesel engine particulate matter. Energy Procedia 68:370–380. https://doi.org/10.1016/j.egypro.2015.03.268

Talebizadeh P, Babaie M, Brown R, Rahimzadeh H, Ristovski Z, Arai M (2014) The role of non-thermal plasma technique in NOx treatment: a review. Renew Sust Energ Rev 40(x):886–901. https://doi.org/10.1016/j.rser.2014.07.194

Thiyagarajan S., Geo VE, Martin LJ, Nagalingam B (2018a) Combined effect of fuel-design and after-treatment system on reduction of local and global emissions from CI engine. Environmental Technology (United Kingdom) 1–11. https://doi.org/10.1080/09593330.2018.1453871

Thiyagarajan S, Varuvel EG, Martin LJ, Beddhannan N, Geo VE, Martin LJ, Nagalingam B (2018b) Mitigation of carbon footprints through a

blend of biofuels and oxygenates, combined with post-combustion capture system in a single cylinder CI engine. Renew Energy 130: 1067–1081. https://doi.org/10.1016/j.renene.2018.07.010

Thiyagarajan S, Sonthalia A, Edwin Geo V, Ashok B, Nanthagopal K, Karthickeyan V, Dhinesh B (2019) Effect of electromagnet-based fuel-reforming system on high-viscous and low-viscous biofuel fueled in heavy-duty CI engine. J Therm Anal Calorim 138(1): 633–644. https://doi.org/10.1007/s10973-019-08123-w

Thomas SE, Shawcross JT, Gillespie R, Raybone D, Martin AR (2001) The role of NO selective catalysts in the plasma enhanced removal of NOx and PM from diesel exhaust. Sae Technical Paper Series 724:2001–3568. https://doi.org/10.4271/2001-01-3568

Thonglek N, Dechthummarong C, Kiatsiriroat T (2011) Soot treatment by using high voltage pulse energized electrostatic precipitator. Energy Procedia 9:292–298. https://doi.org/10.1016/J.EGYPRO.2011.09.031

Tonkyn RG, Barlow SE, Hoard JW (2003) Reduction of NOx in synthetic diesel exhaust via two-step plasma-catalysis treatment. Appl Catal B Environ 40(3):207–217. https://doi.org/10.1016/S0926-3373(02)00150-9

Tran DN, Aardahl CL, Rappe KG, Park PW, Boyer CL (2004) Reduction of NOx by plasma-facilitated catalysis over In-doped γ-alumina. Appl Catal B Environ 48(2):155–164. https://doi.org/10.1016/j.apcatb.2003.10.008

Ueda M, Itoh Y, Shinjoh H, Nakakita K, Arakawa M (2004) A concept of plasma assisted catalyst system using a DeNOx catalyst for an automobile diesel engine. SAE Technical Papers 724. https://doi.org/10.4271/2004-01-1834

Umemoto H, Hayashi H, Takashima K, Mizuno A (2008) Electrostatic precipitators for cleaning diesel exhaust. In 2008 IEEE Industry Applications Society Annual Meeting (pp. 1–4). https://doi.org/10.1109/08IAS.2008.97

Varuvel EG, Sonthalia A, Subramanian T, Aloui F (2018a) NOx-smoke trade-off characteristics of minor vegetable oil blends synergy with oxygenate in a commercial CI engine. Environ Sci Pollut Res 25(35):35715–35724. https://doi.org/10.1007/s11356-018-3484-y

Varuvel EG, Sonthalia A, Subramanian T, Aloui F (2018b) NOx-smoke tradeoff characteristics of minor vegetable oil blends synergy with oxygenate in a commercial CI engine. Environ Sci Pollut Res 25(35):35715–35724

Zukeran A, Ikeda Y, Ehara Y, Ito T, Takahashi T, Kawakami H, Takamatsu T (2000) Agglomeration of particles by ac corona discharge. Electrical Engineering in Japan 130(1):30–37. https://doi.org/10.1002/(SICI)1520-6416(200001)130:1<30::AID-EEJ5>3.0.CO;2-F

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Biofuels
Vol. 10, Issue 2, July-December, 2019 pp. 79-86

IndianJournals.com
A product of Diva Enterprises Pvt. Ltd.
DOI : 10.5958/0976-4763.2019.00011.4

Effect of Using Waste Plastic Oil on the Performance and Emissions Characteristics of Unmodified Constant Speed Direct Injection Diesel Engine

Parvesh Kumar[1,2]*, P. Srinivasulu[1] and Sidharth[2, 3]

[1]Department of Mechanical Engineering, Vaagdevi College of Engineering, Warangal-506005, Telangana, India
[2]Centre for Advanced Studies and Research in Automotive Engineering, Delhi Technological University, Delhi-110042
[3]Department of Mechanical & Automation Engineering, Maharaja Agrasen Institute of Technology, Delhi-110086, India
**Corresponding author email id: pkkhatkar12@gmail.com*

Received: 17-05-2019; Accepted: 09-12-2019

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 14.139.251.100 on dated 29-Jan-2020

ABSTRACT

Waste plastic fuel is a sustainable solution for the biggest dual problem of the world i.e. depilation of fossil fuels and environmental degradation due to waste plastic. In this research the waste plastic is converted into liquid fuel and the fuel is tested on the direct injection CI engine. Before testing on the engine the physico-chemical properties of the WPO are tested and the results are compared with standard petroleum diesel fuel. The performance and emissions characteristics of DI-CI engine are also analysed by using WPO as fuel and the results are compared with conventional diesel fuel. The engine shows better performance at higher load while operated with WPO. The emissions are also better than conventional diesel fuel.

Keywords: WPO, Performance, Pyrolysis, Emissions, Waste Plastic

1. INTRODUCTION

Presently world is facing two biggest problems of economical energy source which is the back born of development and continuously increasing air pollution and global warming [1]. Most of the energy requirement is fulfil by fossil fuels which are available in limited quantity. Also the combustion of fossil fuels release harmful emissions which cause air pollution and global warming [2]. Apart from it, the air pollution is also increase due to waste plastic which is available in huge amount throughout the world. However, the production of plastic on a large scale started nearly 1950 [3]. But, now a days it seems impossible to imagine a world without plastic as it become the necessity of today because of its properties. The word plastic comes from one greek word "plastikos" which means the material which can mould easily [3]. The plastic is invented by a british engineer Alexander Parkes in 1857 which is presented in 1962 at London International Exhibition [4]. However, world first synthetic plastic was invented in 1907 in New York known as Bakelite. Very soon plastic get popularity due to its properties like water proof, durable, light weight, reliable, strength and easily moldable. One property i.e. durability become the biggest problem in usage of plastic.

According to the survey conducted worldwide in 2013 approximately 299 million tons of plastic produced worldwide in 2012 which was supposed to be increased by 4% every year due to increase in population and demand of the plastic [5]. It is found that approximately 141 million tons of plastic used for packaging was disposed or littered worldwide in 2015 which is 86% of total plastic used for packaging purpose [6]. In this wastage 27% is contributed by China and 10% of total waste is contributed by Indonesia [6]. From the study it is noticed that more than 60% of total plastic is

disposed or littered which is a very big amount [5]. According to one report India produce 9.6 million tons of plastic every year and 43% of that used for packaging only. Almost 60% of total plastic remains uncollected which act as a waste to the environment and results in problems [7].

The plastic comes in polymer group which is mainly non-biodegradable. However, plastic mostly contain carbon and hydrogen which are coupled with each other with the help of some other molecules [8]. The carbon and hydrogen atoms are strongly packed in long chain which make it non- biodegradable. Therefore, single use plastic become the problem worldwide [6]. On the other hand as it contain carbon and hydrogen researchers find scope of energy source. Hence, researchers worked to crack the plastics and convert into fuel [9,10]. Pyrolysis found one of the simple, suitable, efficient and economic methods of converting waste plastic into liquid and gaseous fuel. Pyrolysis is the process of breaking of long chain of polymers and plastics by heating and in the absence of oxygen to the shorter chain. Most of plastic breakdown easily by heating in the absence of oxygen [3,5,11]. However, some plastic required some catalysts for breakdown. Catalysts make the process easily and fast also it improves the productivity and quality of the final product[5,9,12, 13].

Three products are received while plastic is heated in the absence of oxygen i.e. non-condensable gases, condensable gases, and solid residue. The non-condensable gases which does not condensed after cooling can be used as gaseous fuel for the thermal cracking process. The quantity of non-condensable gases varies from 5-15% and the quantity depend on the type of plastic and process of cracking waste [3]. The condensable gases get condensed after cooling and it can be used as liquid fuel. The properties of liquid fuel collected from the cracking of waste plastic are comparable to petroleum diesel fuel. The liquid fuel contributes the maximum portion in the cracking process. The portion of liquid fuel collected after cracking varies from 80-95% [10]. Some solid residue left behind after pyrolysis which varies between 2-5% and known as char [5].

In the present work the waste plastic collected from the canteen of Vaagdevi College of Engineering, Warangal is converted into fuel. The fuel collected from the pyrolysis process is tested on constant speed single cylinder four stroke direct injection diesel engine in the lab. The physico-chemical properties of waste plastic fuel is measured and found that most of the properties are comparable to petroleum diesel fuel and some are even better than petroleum diesel fuel. The waste plastic fuel is than tested on diesel engine and the performance and emissions results are compared with conventional diesel engine. The performance of waste plastic fuel is found better than petroleum diesel. Also the emissions like NO_x and CO are less for waste plastic fuel as compared to petroleum diesel fuel.

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 14.139.251.100 on dated 29-Jan-2020

2. EXPERIMENTAL

2.1 Materials

Waste plastic in the form of used water bottles, cold drinks bottles, food packets are collected from the canteen of Vaagdevi College of Engineering (VCE), Warangal. The collected waste plastic is then cut into small pieces. The thickness of plastic collected varies from 40μm to 75μm and the width of cut pieces is not more than 10mm. After cutting the plastic is washed thoroughly with water and after that it put for drying in the sunlight. It took nearly 2-3 days to completely dry.

2.2 Pyrolysis Setup and Procedure

The pyrolysis process is conducted in the Mechanical Engineering Department of VCE, Warangal. The pyrolysis setup consist of one electric furnace of rating 2kW which is used to melt and boil the waste plastic present in glass reactor. Nearly 70% of glass reactor is filled with waste plastic pieces and put on the electric furnace. The glass reactor has three necks. One neck is reserved for feeding of waste plastic pieces and chemicals. The second neck is reserved for

thermocouple which is inserted inside the glass reactor to monitor the inside reactor temperature. Both these two necks are tightly closed with the help of cork. The third neck is reserved for collection of vapour produce during pyrolysis process. K-type thermocouple is used in the present research having range of 1700°C with an accuracy of ±2°C. The thermocouple is connected with digital temperature indicator which shows the reactor temperature. The supply to the furnace is controlled with the help of AC transformer. The vapours of plastic are collected through a pipe which is connected to the condenser. One condenser is used to condense the vapour of plastic collected from the reactor. One chiller is also used for the condenser.

First of all the glass reactor is filled with dry waste plastic small pieces up to 70% of its volume. Turn on the electric furnace and put glass reactor over the furnace. Install thermocouple in one neck of the reactor and vapour collecting tube on the second neck of the reactor. Connect thermocouple with the temperature indicator and controller. Keep the needle valve closed placed on the vapour collecting tube. Before closing the third neck of reactor with cork remove all the air present in the reactor with the help of vacuum cleaner. Connect the vapour connecting tube with condenser in which the waste plastic vapours get condensed and converted into liquid fuel. To remove heat from the waste plastic vapour continuously in the condenser, one chiller is coupled with condenser which supply the cold water to the condenser. In the starting the temperature inside the reactor start increasing and reaches to 160°C. At this temperature the plastic start melting. The temperature of the reactor remains constant at 160°C for some time until unless all plastic melted down. Once all plastic melts, the temperature inside the reactor again starts increasing. After achieving sufficient temperature inside the reactor the plastic starts boiling. Now plastic starts converting into vapours and these vapours moving towards vapour passing tube. Open the needle valve and let the vapour passes through the condenser. In condenser most of the vapour get condensed and converted into liquid which is collected in the beaker. However, some portion of the waste plastic vapours remain uncondensed and goes to the environment. The schematic diagram of the waste plastic fuel conversion unit is shown in Figure 1.

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 14.139.251.100 on dated 29-Jan-2020

2.3 Engine Test Rig and Testing Method

The engine selected for present research is a kirloskar four stroke, single cylinder, vertical, air cooled, manually cracked started, gravity fuel feeding system direct injection diesel engine which is coupled with eddy current dynamometer. The engine has a rated power of 7.5hp (5.9kW) and run at constant speed. The engine is provided with all necessary equipment to measure brake specific fuel consumption, brake thermal efficiency, exhaust gas temperature. The speed of the engine is measured with the help of tachometer. The load on the engine are applied through eddy current dynamometer. The mass flow rate of the fuels is measured by stopwatch and burette which is installed on the

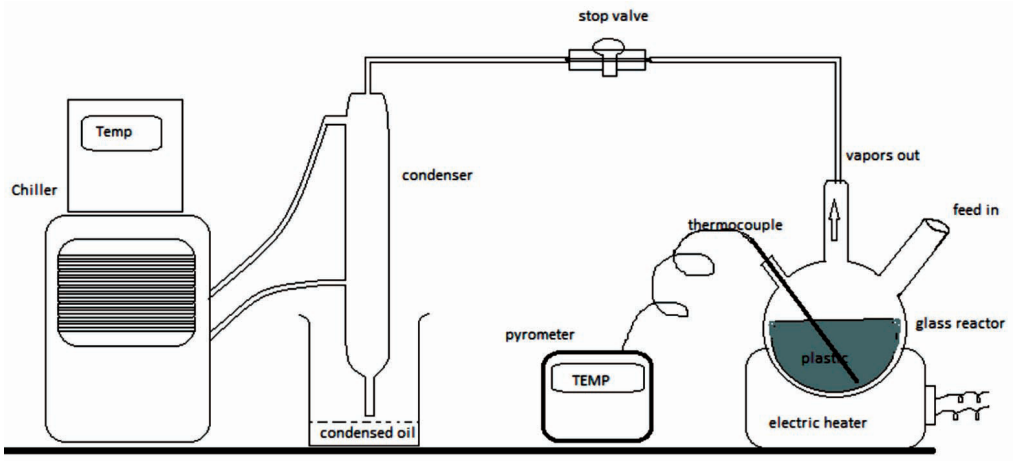


Figure 1: Setup for Pyrolysis and Condensation

control panel of the test rig. The exhaust emissions of the engine like HC, CO and NO_x are measured by AVL 4000 light di gas analyser. However, the AVL 437 smoke meter is used to measure the smoke opacity of the engine. The schematic diagram of the test rig is given in Figure 2 and specifications of the engine is given in Table 1.

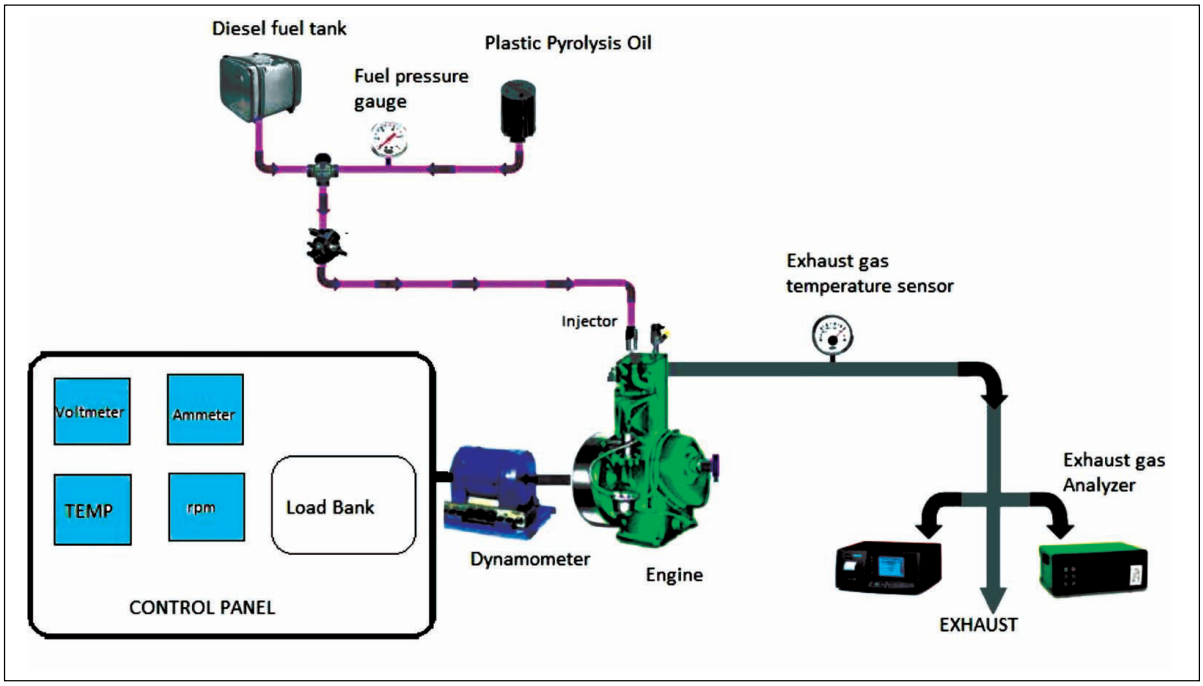


Figure 2: Schematic Diagram of Experimental Test rig

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 14.139.251.100 on dated 29-Jan-2020

Table 1: Technical Specification of the Diesel Engine

Make	**Kirloskar**
Number of cylinder	1
Bore x stroke	95 x 110mm
Cubic capacity	0.78 Lit.
Compression ratio	17.5:1
Rated output as per BS514/ISO 3046/ISO 10001	5.9 KW at 1500 rpm
Starting	Hand start with cranking handle
SFC at rated hp/1500 rpm	251 g/KWh
Fuel tank capacity	11.5 lit.
Fuel tank refilling time period	Every 6.9 hours engine running at rated o/p
Engine weight w/o flywheel	118 kg
Weight of flywheel	64 kg
Rotation while looking at flywheel	Clockwise
Cooling system	Air cooled
Dynamometer	Eddy current

3. RESULT AND DISCUSSIONS

3.1 Physico-chemical Properties

In this section the physico-chemical properties like density, kinematic viscosity, flash point, calorific value, cetane index etc. are measured and the results are compared with petroleum diesel fuel. The most of the properties are measured according to Kumar and Kumar [1]. The values of measured properties in the present research is shown in Table 2.

Table 2: Comparison of Some Properties Diesel, Waste Plastic Oil and Gasoline

Properties	WPO	Diesel
Colour	Pale yellow	Orange
Specific gravity at 30°C	0.8355	0.84 to 0.88
Gross calorific value (kj/kg)	45340	45500
Kinematic viscosity, cSt @ 40°C	1.6	3.1
Cetane number	56	54
Sulphur content (%)	<0.002	<0.035
Flash point °C	42	50
Fire point °C	45	56
Pour point °C	<-7	6

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 14.139.251.100 on dated 29-Jan-2020

It found for the result that some physico-chemical properties like specific gravity or density, cetane number, kinematic viscosity, flash point and fire point of the waste plastic oil (WPO) is better than petroleum diesel fuel. However, the calorific value of WPO is very comparable to petroleum diesel fuel.

3.2 Performance Characteristics

In this section the performance characteristics of diesel engine using WPO as fuel is measured and the results are compared with petroleum diesel fuel. The brake thermal efficiency (BTE), brake specific fuel consumption and exhaust gas temperature of the WPO are measured and the results are compared with conventional diesel engine. The variations in the brake thermal efficiency (BTE), brake specific energy consumption (BSEC) and exhaust gas temperature with respect to (w.r.t.) of WPO and petroleum diesel fuel are shown in Figure 3, 4&5, respectively.

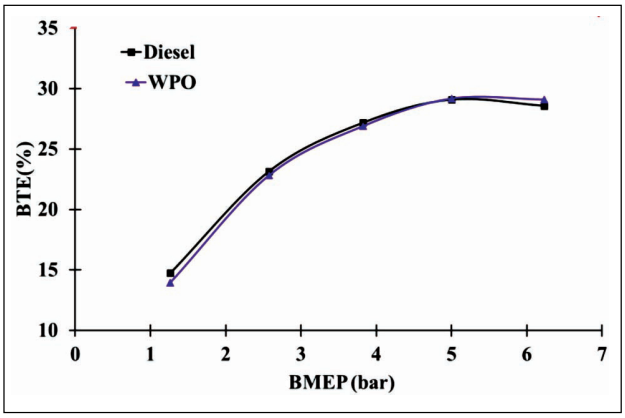


Figure 3: BTE w.r.t. BMEP

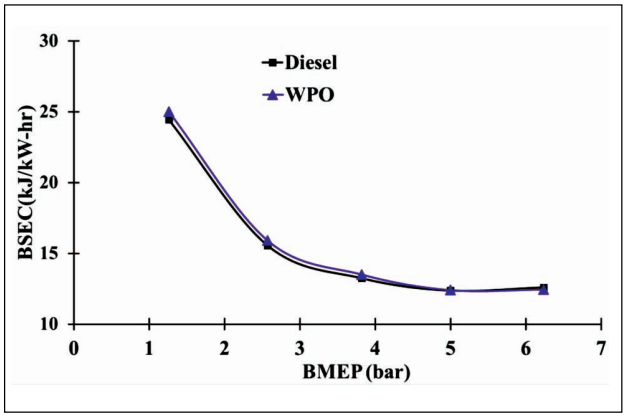


Figure 4: BSEC w.r.t. BMEP

It can be seen from the figure that the performance of direct injection diesel engine is better at higher load when it is running on WPO as compared to running on conventional diesel fuel. However, at lower load the performance of conventional diesel fuel is better as compared to WPO in the same engine. The better performance of WPO over conventional diesel is due to better physico-chemical properties.

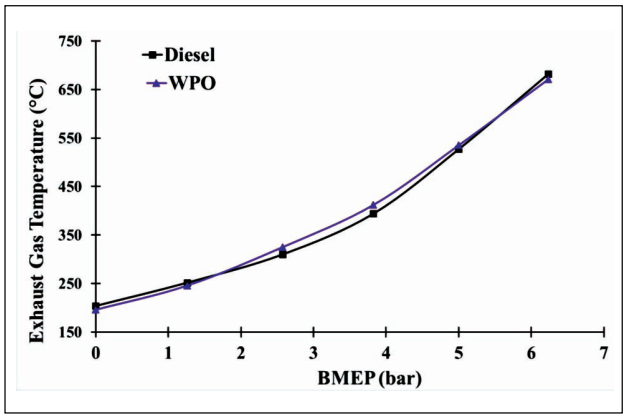


Figure 5: Exhaust gar temperature w.r.t. BMEP

3.3 Exhaust Emissions Characteristics

The exhaust emissions of direct injection diesel engine using WPO as fuel are analysed in this section and the results are compared with petroleum diesel fuel. The emissions like CO, UH, NO_X and smoke opacity are measured and the result is analysed in this section. The variation of CO, HC and NOX emissions w.r.t. BMEP for compressed ignition (CI) engine using WPO and petroleum diesel as fuel is shown in Figure 6, 7 and 8, respectively.

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 14.139.251.100 on dated 29-Jan-2020

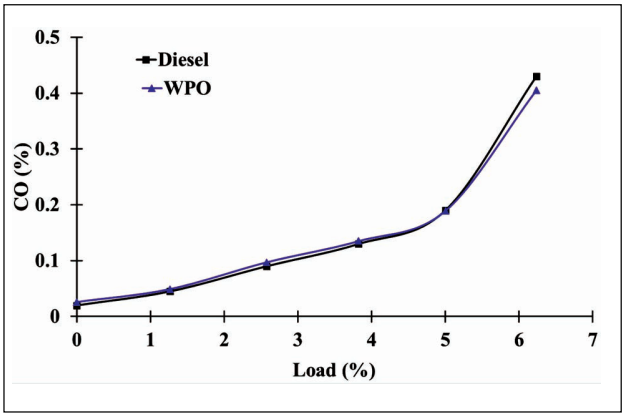


Figure 6: CO emissions w.r.t. BMEP

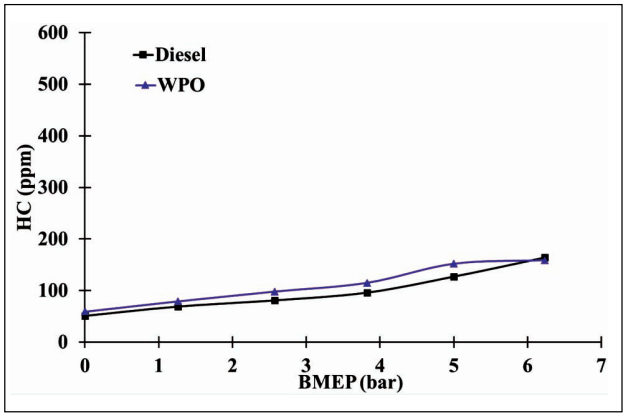


Figure 7: HC emissions w.r.t. BMEP

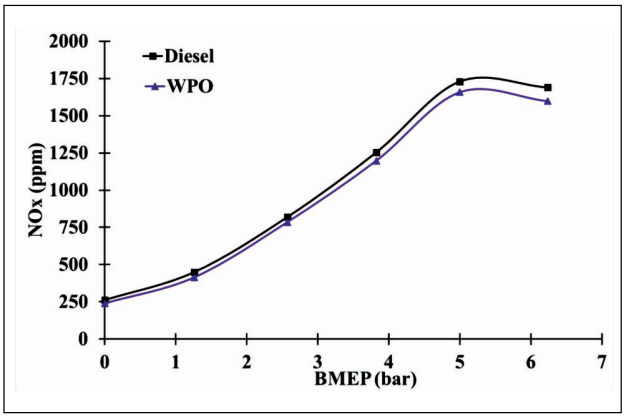


Figure 8: NO_X emissions w.r.t. BMEP

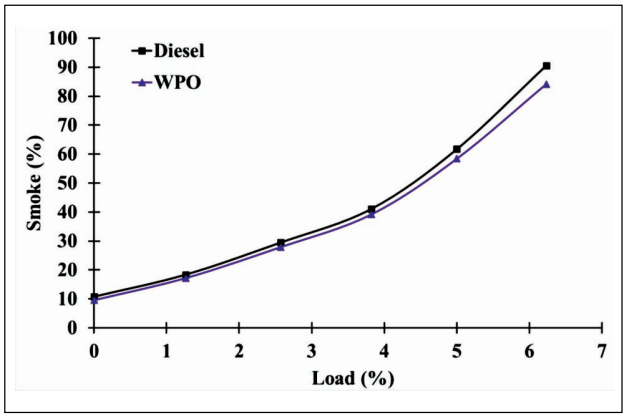


Figure 9: Smoke Opacity w.r.t. BMEP

The CO emissions of WPO fuel are slightly higher than petroleum diesel fuel at low and intermediate loading however the CO emissions decreases at high and full load. At full load the CO emissions of WPO fuels are less than petroleum diesel fuel. On the other hand the HC emissions or WPO fuel are slightly higher than conventional diesel fuel for all loading conditions. The NO_X emissions are very important to decrease due to continuously increasing air pollution. Therefore, the NO_X emissions of WPO fuels are less than petroleum diesel fuel for all loading conditions. This happens due to smooth and low temperature combustion of WPO in the combustion chamber. The smoke opacity of diesel engine using different fuels with respect to BMEP is shown in figure 9. It is clear from the figure that, the smoke opacity of diesel engine is less when WPO is used as fuel as compared to petroleum diesel as fuel.

4. CONCLUSIONS

In the present research the WPO fuel is used as fuel in direct injection CI engine. The performance and emissions characteristics of the engine analysed using WPO and petroleum diesel as fuel. Also the physico-chemical properties of WPO is measured and the results are compared with petroleum diesel fuel. The major finding from the present research are:

- Most of physico-chemical properties of WPO are better than conventional diesel fuel. However, the calorific value of WPO is slightly less than petroleum diesel fuel.
- The performance of diesel engine is better at higher load when running on WPO as compared to simple diesel fuel.
- The NO_X emissions and smoke opacity of diesel engine is less when operated with WPO as fuel as compared to normal diesel for all loading conditions.
- The diesel engine emits slightly higher CO and HC emissions while running of WPO as compared to conventional diesel fuel at low and intermediate loading. However, the emission performance improves at higher load.

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 14.139.251.100 on dated 29-Jan-2020

REFERENCES

[1] Kumar P, Kumar N. Environmental effects process optimization for production of biodiesel from orange peel oil using response surface methodology. Energy Sources, Part A Recovery, Utilization, and Environmental Effects **2019**; 12: 1–11.

[2] Saravanan AP, Mathimani T, Deviram G, Rajendran K, Pugazhendhi A. Biofuel policy in India: a review of policy barriers in sustainable marketing of biofuel. Journal of Cleaner Production **2018**; 193: 734–747.

[3] Tulashie SK, Boadu EK, Dapaah S. Plastic waste to fuel via pyrolysis: A key way to solving the severe plastic waste problem in Ghana. Thermal Science and Engineering Progress **2019**; 11(1): 417–424.

[4] Office UP. Patents for inventions. UK Patent Office; 1857. 255 p. https://www.hup.harvard.edu/catalog.php?isbn=9780674657755

[5] Eswaramoorthi M, Venkateshan T, Bala M, Gejendhiran S. Review of plastic waste management by pyrolysis. International Journal of Research Applied Science Engineering and Technology **2016**; 4(11): 514–517.

[6] Giacovelli C. Single-Use Plastic: A Roadmap for Sustainability. United Nation Environment Programme **2018**; 1–90 p. http://wedocs.unep.org/handle/20.500.11822/25496

[7] India generates 9.46 mn tonnes of plastic waste annually - The Hindu BusinessLine. https://www.thehindubusinessline.com/news/science/india-generates-946-mn-tonnes-of-plastic-waste-annually/article29299108.ece

[8] Arena U, Zaccariello L, Mastellone ML. Fluidized bed gasification of waste-derived fuels. Waste Management **2010**; 30(7): 1212–1219.

[9] Shadangi KP, Mohanty K. Thermal and catalytic pyrolysis of Karanja seed to produce liquid fuel. Fuel **2014**; 115(1): 434–442.

[10] Quesada L, Pérez A, Godoy V, Peula FJ, Calero M, Blázquez G. Optimization of the pyrolysis process of a plastic waste to obtain a liquid fuel using different mathematical models. Energy Conversion and Management **2019**; 188(1): 19–26.

[11] Damodharan D, Sathiyagnanam AP, Rana D, Rajesh Kumar B, Saravanan S. Extraction and characterization of waste plastic oil (WPO) with the effect of n-butanol addition on the performance and emissions of a DI diesel engine fueled with WPO/diesel blends. Energy Conservation Management **2017**; 131(2): 117–26.

[12] Xue Y, Johnston P, Bai X. Effect of catalyst contact mode and gas atmosphere during catalytic pyrolysis of waste plastics. Energy Conservation Management, **2017**; 142(1): 441–451.

[13] Hashem FS, Razek TA, Mashout HA. Rubber and plastic wastes as alternative refused fuel in cement industry. Construction and Building Materials **2019**; 212: 275–282.

How to cite this article: Kumar P, Srinivasulu P and Sidharth. Effect of using waste plastic oil on the performance and emissions characteristics of unmodified constant speed direct injection diesel engine. Journal of Biofuels **2019**; 10(2): 79-86.

www.IndianJournals.com
Members Copy, Not for Commercial Sale
Downloaded From IP - 14.139.251.100 on dated 29-Jan-2020

Int J Syst Assur Eng Manag
https://doi.org/10.1007/s13198-019-00943-1

ORIGINAL ARTICLE

Electrical power system transmission quality and power supplier micro grid control functional reliability

L. N. Das[1] · **Sanyam Gupta**[1]

Received: 10 September 2019 / Revised: 10 September 2019
© The Society for Reliability Engineering, Quality and Operations Management (SREQOM), India and The Division of Operation and Maintenance, Lulea University of Technology, Sweden 2020

Abstract Electrical pulse generator or a transformer transports energy from various sources such as wind farm, electromagnetic field through a gas propelled air turbine, transmits currents through the high sky transmission lines. The consumer industry and domestic user multi machines' power system micro and macro grids control functional has facilities for measuring both synchronous and asynchronous transmittance phased flow current pulses. The petroleum refinery located at islands uses powers from both the wind farm energy and gas fueled hydro-air mixture operated turbine's mechanically air energy. Electric generator holding columns stability, wind farms air pressure regularity are some of the power grid supporting analytics helps measuring a grid potential, transmitter capacity, transmitting fault-line determination, fault type recognition, fault prohibition in the DC and AC system. Mathematical graph models with the differential equation solutions stability graphical lines helps a transmission quality and micro grid control functions reliability measure and requisite transmission path activeness or passiveness decision making process. Decisive outcome motivates system passive components repair and be fit for active up to the active components rest or temporary halt periodic time forecast. Supporting vending machine tasks schedule are automatically adjusted for readily actions at any time from the repairing time. In the present paper writer prefers to review the article of reliability decision making process optimization topics for helping the power plant maintenance servers cum reinforcement machine learners.

Keywords Wind farm energy · Gas fueled hydro-air mixture operated mechanical energy · Stability analysis · Transmitting fault line analysis

1 Introduction

Different models are defined to study electrical power system quality and micro macro grid control functions' reliability. The micro smart grids using in aircraft power system, Electronic computing machines power systems quality are reliable for the user as users without any hesitation accepts the norms of micro controller power grids functionality and descriptive picture images.

The paper title "Robust fault analysis in transmission lines using synchrophasor measurements" (Rajaraman et al. 2018) abstracts the methodology for detection, classification and localization of transmission line faults using synchronous phase measurements. The techniques involves the extraction of phasors from the instantaneous three phase voltages and currents at each bus in the system which are then decomposed into their symmetrical components. The current through the symmetric components are sent to the phasor data concentrator (PDO) for real time fault analysis, which is completed within 2–3 cycles after fault inception. Rajaraman et al. (2018) have proposed transmission algorithm independent of symmetric component system topology. Rajaraman et al. (2018) convinced the accuracy and speed of the transmission technique can be communicated appropriately to facilitate system restoration

✉ Sanyam Gupta
sanyam_phd2k18@dtu.ac.in

L. N. Das
lndas@dce.ac.in

[1] Department of Applied Mathematics, Delhi Technological University, New Delhi, Delhi 110042, India

Published online: 04 January 2020

Electrical pulse generator transform power from various sources such as wind farm, electromagnetic field or through a gas fueled air turbine. The electric transformer grid transmits currents through the high sky transmission lines. The consumer industry and domestic user multi machines power system micro and macro grids control functional has facilities for measuring both synchronous and asynchronous transmittance phased flow current pulses. The quality of transformed electric phased flow current pulses are estimated through the plotted graphic lines or meters. The performance of petroleum refinery located at islands uses powers from both the wind farm energy and gas fueled hydro-air mixture operated turbine's mechanically air energy are also measured from analyzing the pictures of the mechanical operator parameters. The cyclone and bad weather periodic operation analysis are also performed by observing the refinery pipe and container holding columns stability, wind farms air pressure regularity etc. These are some of the power grid supporting analytics helps measuring a grid potential, transmitter capacity, transmitting fault-line determination, fault type recognition, fault prohibition in the DC and AC system. Mathematical graph models with the differential equation solutions stability graphical lines help a transmission quality and micro grid control functions reliability measure and requisite transmission path activeness or passiveness decision making process. Outcome of the decision motivates system passive components repair and be fit for active up to the active components rest or temporary halt periodic time forecast. Supporting vending machines tasks schedule are automatically adjusted for readily actions at any time from the repairing time. In the present paper writer prefers to review the article of reliability decision making process chapter of Optimization topics for helping the power plant maintenance servers cum reinforcement learners.

2 Power quality and power system stability

The stability of a power transmittance system is analysed either by single step differential equation model or by probable random variables multistep models. Here we first describe differential equation model than discuss the random variable models.

2.1 Single step differential equation model for stability analysis

Let $f(x, y)$ be a power transmitting system variable. If $\frac{dy}{dx} = f(x,y)$ does not change too rapidly near a space position (x_0, y_0) then $y' = f(x_0,y_0) + (y - y_0)f_y(x_0,y_0) + (x - x_0)f_x(x_0,y_0)\ldots$ Let us simplify the stability equation by writing $y' = \lambda y + ax + b$ here parameters $\lambda = f_y(x_0,y_0), a = f_x(x_0,y_0)$ and $b = f(x_0,y_0) - y_0 f_y(x_0,y_0) - x_0 f_x(x_0,y_0)$. The solution of the differential equation is $y = ae^{\lambda x} + p(x)$ where p(x) is a polynomial of at most degree 2 and the solution is dominated by the exponential term. Also the equation $y' = \lambda y$ tests the partial stability of transmitting system variable. The parameter $\lambda = a + jb$, where $j = \sqrt{(-1)}$, a and b are constants. The system is partially stable for $a < 0$.

Example 2.1 Let the variation of current is represented in the following form, $\frac{dI}{dt} = I^2$, $I(0) = 1$ varying over time interval [0, 1]. Using single step method it can be verified $I(t) = \frac{1}{(1-t)}$. By the Single step method we use the iteration $I_{n+1} = I_n + hf(t_n, I_n)$. Here $f(t, I) = I^2$, assume $h = 0.2, t_n = t_0 + nh$, $n = 1, 2, \ldots, 5, t_0 = 0$ and $I_0 = 1$. Therefore $I_1 = I_0 + 0.2f(t_0, I_0) = 1.2$. Using differential equation solution function $I(t) = \frac{1}{(1-t)}$ we get $I(0.2) = \frac{1}{(1-0.2)} = \frac{1}{0.8} = 1.25$. Repeating the iteration we obtain the following tabular values (Table 1).

Using single step stability analysis we get the equation $I' = \lambda I + at + b$. Here $\lambda = f_I(t_0, I_0) = \frac{dI^2}{dt}_{(0,1)} = 2, a = f_t(t_0, I_0) = \frac{dI^2}{dt}_{(0,1)} = \frac{d(\frac{1}{(1-t)})^2}{dt} = 2$ and $b = f(t_0, I_0) - I_0 f_I(t_0, I_0) - t_0 f_t(t_0, I_0) = 1 - 1 \times 2 - 0 \times 2 = -1$. Substituting the values of λ, a and b in the equation $I' = \lambda I + at + b$, we get $I' = 2I + 2t - 1$. The integrating factor is $I = e^{2t}$. The solution to the linear DE is $\int (2t - 1)e^{2t} dt$. Thus $I(t) = (t - 1)e^{2t}$. If $f(t, I)$ does not change rapidly then the uniformity in current flow occurs.

2.2 Probable random variable multistep method

A multistep method for solving a first order initial value problem representing power transmission through a conductive material is using the last computed points $(x_k, y_k), k = m, m-1, m-2, \ldots$ to compute the next points (x_{m+1}, y_{m+1}) intensity. Some times a pair of multistep methods is used to solve the problem where one can

Table 1 Comparision between numerical values obtain from single step method and the differential equation

n	t_n	I_n by single step method	I_n computed using DE solution
0	0.0	1.000	1.00
1	0.2	1.200	1.250
2	0.4	1.488	1.887
3	0.6	1.931	2.500
4	0.9	2.676	5.000
5	1.00	4.109	

predict the value of y_{m+1} and other is correcting the predicted value of y_{m+1}. A recursive form for solution of $\frac{dy}{dx} = f(x, y)$ for the general linear multistep method with step size m is an equation

$$Y_{n+1} + \alpha_1 Y_n + \alpha_2 Y_{n-1} + \cdots + \alpha_m Y_{n-m+1} = h(\beta_0 f_{n+1} + \beta_1 f_n + \cdots + \beta_m f_{n-m+1}) \quad (1)$$

where $f_n = f(X_n, Y_n)$ and $\alpha_1, \alpha_2, \ldots, \alpha_m, \beta_0, \beta_1, \ldots, \beta_m$ are constants. If $\beta_0 = 0$, then the method is explicit otherwise implicit. When $m = 1$ the method is reduced to single step method. Suppose the differential equation is $y' = \lambda y$ and $\lambda < 0$. Then the multistep form is

$$(1 - h\beta_0\lambda)Y_{n+1} + \sum(\alpha_k - h\lambda\beta_k)Y_{n+1-k} = 0 \quad (2)$$

The solutions of the above recurrence relations are $Y_n = cr^n$, where c is a constant and r is the solution of the following equation

$$p(r) = (1 - h\beta_0\lambda)r^m + \sum(\alpha_k - h\lambda\beta_k)r^{m-k} \quad (3)$$

The difference equation in Eq. (1) are associated with m-th order polynomial of r as shown in Eq. (3). Assume $Q(r) = \sum \alpha_k r^{m-k}$ and $R(r) = \sum \beta_k r^{m-k}$, where $\alpha_0 = 1$. Therefore the Eq. (3) can be written as $p(r) = Q(r) - h\lambda R(r)$.The linear multistep method is called zero stable or stable if all the roots of Q(r) lie within or on the unit circle and the roots on the unit circle are not complex.

3 Power transmission quality

Often Curve fitting algorithms are used to link the multi points to test the graphic line and point assembling quality. In the paper Haider et al. (2018) we find the authors have described a dual control strategy for power sharing improvement in islanded mode of AC microgrid where the stability analysis can be studied. In the paper Kornatowski et al. (2008) a probabilistic measure of colour image processing fidelity is discussed. The authors used probabilistic symbols to assess qualities of proposed image processing algorithm and suggested scalar measure can be used for any colour space and the similar results can be obtained on the image content. We feel the power system picture analysis tools are also uses curve fitting parameter computing monitors. The curve fitting is the process of constructing a curve from discrete values of the mathematical function and best fits to a series of data points, possibly subject to the constraints. Most of the time a curve fit algorithm based on equations that can be used to find points along the curve. The power transmission quality can be observed analyzing air pressure, temperature curves through a telephonic digital electromagnetic signals quality and signal processing phase interlaced techniques are interpreted through the LED pictures sensors. Which is a quality measure as in the normal weather, pictures quality differs from cloudy or stormy weather. Control of electronic radio waves quality, propagating through the electromagnetic medium is a reliable instrumentation of learning the transmission line transmittance accuracy and faulty line diagnosis. The space air pressure, temperature radiations at time dependent field is represented in the form of differential equations. The solution function to the differential equations are represent in the form of curves. The discrete points on the curve are linked with a smooth function interpolated polynomial form. The mathematical statistical measure less emphasizes electromagnetic field effect rather the software hardware interface technician uses both the number assigned statistics those quantify the electromagnetic field strength observing picture qualities.

Efficient electronics computer programmer uses numerical techniques recurrence relation and iteration or loops in programming language to interpolate the discrete functional values to represent points joining lines, arcs, boundaries and covering the gaps in the picture or textures that represent power transmission quality and highlights through the colour images using optoelectronic tools.

The statistics literature uses words curve fitting through regression, mathematics coordinate geometry, numerical analysis describes the above technique as curve tracing and all these formulations are used in the image processing software hardware interlaces. The Optoelectronics instrumentation technique uses lesser ray beams to carve the camera images or scanned signals in image forms within electronic materials surfaces and can be restored and reprocessed at latter times for edit and reproduced.

Computer text editor, photo editor, image process and electronically drafter tools micro processing techniques are governed by electromagnetic wave front signals stability and magnetic colour ink dynamics regulator and batteries. So quality of printing in the LCD, LED screens, photographic light beholder materials are reliable maters.

4 Conclusion

In the paper Sects. 2.1 and 2.2 we discuss the single step and multistep methods for measuring uniformity or stability of current flow in a system. The contents of Sect. 3 is a linguistic representation of current transmission quality and reliability and originally written to present in the

international conference in quality and reliability. The software and hardware manufactured after or before the conference may be softcopies of information for Computer, Healthcare, engineering, technology quality and reliability purpose.

References

Haider S, Li G, Wang K (2018) A dual control strategy for power sharing improvement in islanded mode of AC microgrid. Prot Control Modern Power Syst. https://doi.org/10.1186/s41601-018-0084-2

Kornatowski E et al (2008) Probabilistic measure of colour image processing fidelity. J Electr Eng 59(1):29–33

Rajaraman P, Sundaravaradan NA, Mallikarjuna B, Jaya Bharata Reddy M, Mohanta DK (2018) Robust fault analysis in transmission lines using Synchrophasor measurements. Prot Control Modern Power Syst. https://doi.org/10.1186/s41601-018-0082-4

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Check for updates

RESEARCH

includes research articles that focus on the analysis and resolution of managerial and academic issues based on analytical and empirical or case research

Empowering Women Through Financial Inclusion: A Study of Urban Slum

Shivangi Bhatia and Seema Singh

VIKALPA
The Journal for Decision Makers
44(4) 182–197, 2019
© 2019 Indian Institute of Management, Ahmedabad
Reprints and permissions:
in.sagepub.com/journals-permissions-india
DOI: 10.1177/0256090919897809
journals.sagepub.com/home/vik
SAGE

Executive Summary

Financial Inclusion is considered to be a critical indicator for development and well-being of the society around the globe. Providing inclusive financial services, that is, financial services affordable for all, has become a basic priority in many countries including India. G-20 Nations have emphasized on financial inclusion as a facilitator for achieving gender equality and other sustainable development goals. Women empowerment is a radical approach concerned with transforming power relations in favour of female gender and considered essential for global progress. Thus, an inclusive financial model is being adopted in developing nations to achieve fundamental to formative objectives. The current article investigates the dimensions of women empowerment, that is, social, political, and economic. It also undertakes a test to see if the dimensions change with financial inclusion. The authors draw upon literature to develop a structured questionnaire on women empowerment and financial inclusion through schemes like Pradhan Mantri Jan Dhan Yojana (PMJDY), Pradhan Mantri Jivan Jyoti Bima Yojana (PMJJBY), Pradhan Mantri Suraksha Bima Yojana (PMSBY), and Atal Pension Yojana (APY) on women living in urban slums in the industrial town of Ludhiana, Punjab. The data were collected from 737 females living in urban slums with PMJDY bank accounts. The result indicated that PMJDY scheme has been quite successful especially in case of women in slums and has a positive influence on social, political, and economic dimensions of women empowerment. The study contributes to existing literature by advancing the debate on women in urban slums and identifies the substantial need for the development of formal financial system to enhance the scale of financial inclusion.

KEY WORDS

Financial inclusion
women in slums
PMJDY
APY
PMJJBY
PMSBY

Note: An extended version of the article titled 'State of Financial Inclusion of Women in Slums in Ludhiana: Post PMJDY' was presented at 60th Labour Economics Conference 19th–21st December 2018.

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://www.creativecommons.org/licenses/by-nc/4.0/) which permits non-Commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

There is no tool for development more effective than women empowerment.

—Mr Kofi Annan

In 2005, the former general secretary of the United Nations quoted this statement to address the issue of gender inequality as a barrier in global development. He also emphasized upon the significance of 'Inclusive Financial Sector' to ensure gender economic equality. As per Findex data of Word Bank, 1.2 billion adults have access to financial services since 2011, yet close to one-third, that is, 1.7 billion adults are still unbanked. About half of the unbanked population includes poor households or out of workforce (The World Bank, 2018). The G-20 summit, in 2010, perceived financial inclusion (FI) as one of the paramount pillars of the worldwide progress. In fact, financial inclusion has been identified as facilitator for seven of the 17 sustainable development goals. World Bank considers FI as the key to boost prosperity by poverty reduction. Since 2010, more than 60 countries have launched a national strategy directed towards attaining Universal Financial Access by 2020.

Women empowerment is a radical approach to transform power relations in favour of female gender that leads to better gender equality (Batliwala, 2007). This enables females to make their life choices, which in turn, effectively improves their well-being. Gender equality and women empowerment are essential to global progress and it can be enhanced by providing affordable financial services to women (Holloway, Niazi, & Rouse, 2017). After the persuasions by G-20 in 2014, financial services were perfused to a vast section of the society, and the period between 2011 and 2014 witnessed an upward trend in the number of first-time adults as bank account holders. On the flip side, it was not able to fill the gender gap for access to basic banking services (Ghosh & Vinod, 2017). This led to social exclusion and gender disparity, highly rampant in case of developing country as compared to a developed country (Ahmed, Aurora, Biru, & Salvini, 2001; Dawar & Singh, 2016). In the words of Noble Laureate Amartaya Sen, 'Poverty is not merely lowness of income, but deprivation of basic capabilities' (2014). Thus, accomplishing complete financial inclusion does not just determine the issues identified with financial structure, rather its centre is annulling the condition of social exclusion (Rangarajan Committee, 2008). Thus, the inclusive financial model has emerged as an arrangement in developing nations to achieve formative objectives. Formulation of mechanism to achieve women empowerment through affordable financial services is a rigorous approach to achieve sustainable growth globally.

The concerns for women empowerment have been rising in India over last few decades. Series of political events post-2014 in India has heightened societal concerns about women's role in economic life as well as critical roles within their households. Mostly, empirical articles in this literature have studied the effects of financial inclusion on women empowerment that has evaluated an over-broadened meaning of empowerment or a truncated part of it (Goetz & Gupta, 1996). Most investigations are typically cross-country research (Demirguc Kunt, Klapper, & Singer, 2013; Lampietti & Stalker, 2000; Quisumbing, Haddad, & Pena, 1995). Within a nationalized context, the studies address the conduct of female-headed family units concentrating basically on financial access alone (Fletschner, 2008; Hazarika & Guha-Khasnobis, 2008; Rawlings & Rubio, 2005). Another set of studies implies presence of gender gap due to lower financial literacy (Fernandes, Lynch, & Netemeyer, 2014), behaviour biases (Frisancho, 2016) and institutional segregation (Corsi & De Angelis, 2017). Estimation of women empowerment is another glitch as it cannot be straightforwardly observed and has numerous features (Beteta, 2006; Mason, 2005; Swain & Wallentin, 2009).

In the Indian context, most of the studies (Datta & Singh, 2018; Ghosh & Vinod, 2017; Swamy, 2014) have used publicly available data to determine the extent of FI. Studies using primary data are limited in number. For women, the phenomena of urbanization and the growth of city ghettoes have unique causes and unique consequences, and yet these issues are largely ignored by prior studies in this area (COHRE, 2008). Only a handful of studies specifically focussed upon links between south Asian slums and women empowerment (Fisher, 2008; Hazarika, 2010; Kaur, Singh, Gupta, Bahuguna, & Rani, 2015; Nasrin, 2012). These studies relied solely on reviews of existing literature and evaluations based on secondary data. On the other side, the current study utilizes primary data from urban ghettos. The convincing motivations to examine women empowerment in this study are multifold. First, women represent two-fifth of work power, yet access to formal financial channels is very low. Second, government and RBI, both, have attempted strides to improve the number of financially included women, and its effect is yet to be assessed. Third, India is one of the developing economies for which the household-level information is promptly accessible, on both access and usage of financial inclusion schemes,

by gender orientation. Fourth, most studies in literature have discussed financial inclusion of rural poor to evaluate and formulate strategies so that the involvement of rural poor may be enhanced in economy. However, the state of financial inclusion among the slum dwellers is largely underexplored in the literature. This is useful and relevant because the evaluation of expenditure, saving, and credit pattern among urban poor may help in formulating effective strategies to make them inclusive.

To contribute to literature, the current study investigates the degree of women's financial inclusion in urban ghettos and how it empowers them. To be increasingly explicit, the study inspects the present status of financial inclusion plans of the Government of India for women in ghettos of industrial city, that is, Ludhiana (India) and how it may affect women. In this study, empowerment is estimated based on social, political, and economic empowerment of women in ghettos. The research framework estimates the impact of the financial inclusion schemes like PMJDY, PMJJBY, PMSBY, and APY on women living in urban slums in the industrial town of Ludhiana. The outcomes in the study are robust to demonstrate that there is a huge increment in the women empowerment due to the above-mentioned schemes. To control the extreme rate of deviations in this study, the sample has been drawn from the population of women who possess bank accounts under PMJDY. This study anticipates that the pace of women empowerment is probably going to trend upwards after they begin benefiting with higher financial inclusion schemes.

FINANCIAL INCLUSION IN INDIA

Financial inclusion brings unbanked and under-banked people in the financial system to provide them the opportunity to access the financial services in order to create economic growth and leads to empowerment opportunities (Lenka & Barik, 2018). The Reserve Bank of India (RBI), in 2015, defines financial inclusion as 'the process of ensuring access to appropriate financial products and services needed by vulnerable groups such as weaker sections and low income groups at an affordable cost in a fair and transparent manner by mainstream institutional players.'

Financial exclusion is a concern for developed as well as developing countries (Dymski, 2005). Even a 'well-developed' financial system has not been successful in bringing universal financial inclusion in many countries (Lenka & Barik, 2018).

Financial Inclusion in the Indian context in the post-liberalization period has witnessed a significant increase in the number of bank branches in urban, semi-urban, and metropolitan regions. Yet, a large section of population is deprived of financial products and services in India at present (Chaia, 2009; Fuller, 1998; Peachy & Roe, 2004). To counter this situation, the Government of India has taken important steps such as the launch of PMJDY in 2014 for the 20–65 age group, with an aim to make financial services, such as banking, insurance, and pension, accessible at an affordable rate for lower- and middle-income groups. These schemes especially targeted women for their financial security (Table 1). These schemes have been idealized to keep with the pace of economic growth, and further to meet sustainable development goals (SDG). Interestingly, there has been an increased expectation on their effect on empowerment of women. Currently, 53 per cent women own bank accounts (Table 2) and majority of them are associated with public sector banks as on May 2019.

Table 1: Financial Inclusion Schemes Launched by the Government of India

Scheme	Details and Benefits	Eligibility
PMJDY launched in 2014	• Accidental insurance cover of ₹ 0.2 million • Life cover of ₹ 30,000 • Overdraft facility up to ₹ 10,000 to preferably lady of the household • RuPay ATM-cum-Debit Card	• All citizen of India with valid proofs • Preference is given to female of household
PMJJBY launched in 2015	• Pay premium is ₹ 330 per annum • Risk coverage under this scheme is for ₹ 0.2 million in case of death	• People in the age group of 18–50 years • Having bank account under PMJDY
PMSBY launched in 2015	• Pay premium of ₹ 12 per annum • The risk coverage under the scheme is ₹ 0.2 million for accidental death and full disability and ₹ 0.1 million for partial disability	• People in the age group 18–70 years • Having bank account under PMJDY
APY launched in 2015	• Fixed pension for the subscribers ranging between ₹ 1,000 to ₹ 5,000	• Minimum age of joining APY is 18 years and maximum age is 40 years • Having bank account under PMJDY

Source: Financialservices.gov.in

Table 2: Accounts Under PMJDY (All figures in ₹ million)

Bank Type	Beneficiaries at Rural/ Semi-Urban centre Bank Branches	Beneficiaries at Urban Metro Centre Bank Branches	Rural-Urban Female Beneficiaries	Total Beneficiaries	Deposits in Accounts	Ru pay Debit Cards Issued to Beneficiaries
PSB	153.4	129.7	149	283.1	781,556.7	223.7
RRB	51.2	9.8	34	61	172,782.7	38.5
PRB	6.9	5.5	6.7	12.4	29,806.1	11.6
Total	211.5	145.1	189.7	356.5	984,145.5	2774

Source: PMJDY Progress Report, GOI, May 2019. Retrieved from https://www.pmjdy.gov.in/account

Notes: PSB—Public Sector Banks; RRB—Regional Rural Banks; PRB—Private Sector Banks.

CURRENT STATE OF RESEARCH ON FINANCIAL INCLUSION

Majority of the literature has focussed on the theoretical and conceptual development of financial inclusion in contexts of micro-finance institutes and self-help groups. An immense assemblage of literature can be found on financial inclusion in India for instance, self-help groups (Basu & Srivastava, 2005; Shah, Rao, & Shankar, 2007; Sinha, 2006; Swain & Varghese, 2009), and loans without collateral by the *Grameen banks (Dev, 2006; Yunus, 2004), and current conditions of financial inclusion schemes that propelled after 2014 (Pillai, 2016; Singh & Naik, 2018; Tulasi, Golait, Sethi, & Goel, 2017; Verma & Garg, 2016).*

A significant number of studies have measured financial inclusion with secondary data sets to identify the levels of financial inclusion in different regions using multidimensional index of banking services (Ambarkhane, Singh, & Venkataramani, 2014; Kim, 2016; Laha, 2015; Laha & Kuri, 2014; Park & Mercado, 2016; Raichoudhury, 2016). The basis of index construction is 'distance-based approach' used by UNDP. Chakravarty and Pal (2013) constructed financial inclusion index through axiomatic method covering data of Indian States from 1972 to 2009. Certain financial inclusion indexes were created using principal component analysis method (Bagil & Dutta, 2012; Lenka & Bairwa, 2016; Lenka & Sharma, 2017). Some of the studies (Arora, 2010; Bihari, 2011; Kumar & Mishra, 2009) created financial inclusion index based on financial access, availability, usage of formal financial services, demand side, and supply side indicators.

The research using primary data set is limited. A few studies attempted to identify the crucial factors for the extent of financial inclusion in India, that is, education, income, financial information, access, usage, and self-help groups (Bhanot, Bapat, & Bera, 2012; Bhutoria & Vignoles, 2018; Siddiqui & Siddiqui, 2017). Thus, the variable 'financial inclusion' is measured along these lines, that is, accessibility and ease of use of financial services for various group of individuals. The current data set in India shows that the number of bank account holders has risen to include around 80 per cent of adults under formal financial framework (Global Findex Report, 2017). As per RBI, 251 million account holders use mobile banking in 2018 and volume of transactions through mobile banking has crossed 3,047 billion in 2018–2019. The ongoing financial inclusion plans, such as PMJDY, PMSBY, and PMJJBY, have encouraged poor and marginalized Indians to open zero-parity ledger and avail insurance protection schemes at affordable rates (Barik & Sharma, 2019). These schemes have helped to top off financial inclusion gaps among various gathering of populace with 55 per cent expansion in women citizens' share. Recently, Bapat and Bhattacharya (2016) investigated changes in financial inclusion of urban poor through a qualitative survey. Using a sample data of 202 slum dwellers in Pune (India), the study indicated that nuclear families with younger age groups and higher willingness-to-save have positive influence on financial inclusion of urban poor. These families have keen interest in availing loans from formal financial institutions.

Despite the fact that substantial advancement has occurred towards financial inclusion in India, a large number of the accounts have been underutilized by the bank account holders. According to Global Findex database, only 20 per cent of Indian adults have active savings bank accounts. The circumstances concerning access of credit from formal financial establishments are even more regrettable as only 7 per cent of Indian adults can access credit from their accounts. Significant reasons for this exclusion from the financial framework are low financial literacy (Fernandes et al., 2014; Lusardi & Mitchell, 2014), social inclinations (Frisancho, 2016; Karlan, McConnell, Mullainathan, & Zinman, 2016), institutional biases (Agier & Szafarz, 2013; Brana, 2012; Corsi & De Angelis, 2017; Fletschner, 2009;

Muravyev, Talavera, & Schäfer, 2009), social limitation (Bylander, 2014; Guérin, Kumar, & Agier, 2013; Hummel, 2013), and absence of trust on financial organizations (Karlan, Ratan, & Zinman, 2014; Mehrotra et al., 2016).

EMPOWERMENT DIMENSIONS

Empowerment is defined as 'Multi-dimensional social process that helps people gain control over their own lives. It is a process that fosters power in people, for use in their own lives, their communities, and in their society, by acting on issues that they define as important' (Page & Czuba, 1999). Of the 17 objectives of United Nation Sustainable Development Goals (SDG) which are considered imperative for worldwide congruity and harmony, objective # 5 for gender balance and women empowerment is viewed as significant for transforming the world through giving equal chance to both of the genders. There are significant empirical evidences recommending that gender imbalance, to a great extent, harm economic advancement (Klasen, 2002), and gender difference in employment and education prompts negative effect on nation's long stint development (Klasen & Lamanna, 2009).

Kabeer (1999) has defined women empowerment as 'the ability to make strategic life choice in context where this ability was previously denied to them'. Women empowerment has received increased scholarly attention in last several decades as an avenue of enhancing mobility, health, economic condition, and participation in decision-making by women (Afrin, 2008). 'Empowerment' is a multidimensional concept and illustrates proliferation of end results. Owning bank account improves financial well-being of women (Ashraf, Karlan, & Yin, 2010; Swamy, 2014).

Political Empowerment

Gender equality is the core to human rights and a bedrock for achieving sustainable development. Increasing women participation in political leadership and decision-making is essential for economic and social development of economy (United Nations, 2018). Thus, political factors such as political quotas, awareness, and voting rights are important interventions in this regard (Armendáriz & Morduch, 2010; Deininger & Liu, 2013; Kabeer, 1999; Sundström, Paxton, Wang, & Lindberg, 2017). Administrative framework, social standards, independence in basic leadership, social versatility and systems additionally significantly affect women empowerment procedures (Bandiera et al., 2018; Beteta, 2006; Chatterjee, Gupta, & Upadhyay, 2018; Kumar et. al., 2019). Political empowerment is advancement of political inclusivity and supporting political involvement as per rights and legitimate job of women. Political empowerment can be accomplished through political awareness, position of power, support in political action, and participation in political groups (Chatterjee et al., 2018). Expanding female portrayal in administration, casting a ballot, and political contribution and inclusion in casual groups to straighten out network issues leads further women empowerment (Bardhan & Klasen, 1999).

Economic Empowerment

Economic empowerment is the capacity to contribute towards the growth processes in a way that recognizes the value of their contributions and makes a fair distribution of their wealth to enhance the access of economic resources (OECD, 2011). Swain and Wallentin (2009) found that economic factor contributes significantly to women empowerment. In another study, Al-Mamun, Wahab, Mazumder, and Su (2014) investigated the change in women empowerment through qualitative survey that quantified economic empowerment utilizing following criteria: decision-making at household level, financial/economic security, resource control, say/control in familial decisions, legal awareness, and mobility. Based on the sample data of 242 low-income urban women in Malaysia taking interest in micro-credit programme, the study quantified the effects of household basic leadership. In an extensive study, Chiapa, Prina, and Parker (2016), through field investigations of female heads of households reviewing 1,236 families in urban ghettos, found that access to savings bank accounts help poor families to handle their assets better, to prioritize education and consumption expenditure, and to feel more responsible for their financial/economic circumstances.

Social Empowerment

Social empowerment means giving power or authority to an individual to improve their livelihood. Person-to-person communication and correspondence, and more noteworthy, social versatility, help in empowerment of women (Swain & Wallentin, 2009). Searing and Chiappori (1998) recommended that social pressure may likewise adjust the women's capacity inside family unit for decision-making. Economic empowerment incorporates economic independence, budgetary self-adequacy, and financial literacy that leads to learning

and aptitudes to deal with financial prosperity (Postmus, 2010; Postmus, Plummer, McMahon, & Zurlo, 2013). Nonetheless, all the dimensions of empowerment, such as social, political, and economic empowerment, are all well-integrated and cannot be secluded (Taylor, 2000).

The relationship between financial inclusion and gender inequality has been underexplored and needs further investigation. Several studies (i.e., Fraser, 2011; Karim, 2011; Roodman and Qureshi, 2006) have raised questionable record on financial inclusion and women empowerment. For example, Bateman (2012) has criticized financial inclusion as 'emerging new label' and considered it mere rebranding of microfinance. Also, financial inclusion drives over indebtedness and lacks demonstrable poverty impact (Duvendack et al., 2011; Guérin, Labie, & Servet, 2015; Stewart et. al., 2012). However, significant number of studies claimed that having bank account can improve the financial well-being of women (Ashraf et al., 2010; Swamy, 2014).

Thus, the proposed hypothesis is as follows:

> **H1:** levels of financial inclusion positively influence different dimensions of empowerment—social, political, and economic.

The framework of the current study is shown in Figure 1, that is, financial inclusion leads to women empowerment. There is tremendous literature on financial inclusion and its capacity in impacting women empowerment. But few studies have examined the role of financial inclusion in empowering female ghetto inhabitants. This is important because there is far too much vulnerability in the lives of the poor and those just living above poverty line. Around 93 per cent of labour forces work in the informal sector. Vulnerability with respect to women managing finance at home rises due to the lack of access to formal financial services. This study attempts to examine financial inclusion of women in urban ghettos of Ludhiana post-PMJDY. Punjab is the fifth most urbanized state in India. With 133 urban centres recently included in Ludhiana city, it has the highest urban population in Punjab. Ludhiana (Punjab) is a prime industrial centre in northern India and is centre point for the hosiery industry and other small-scale industries.

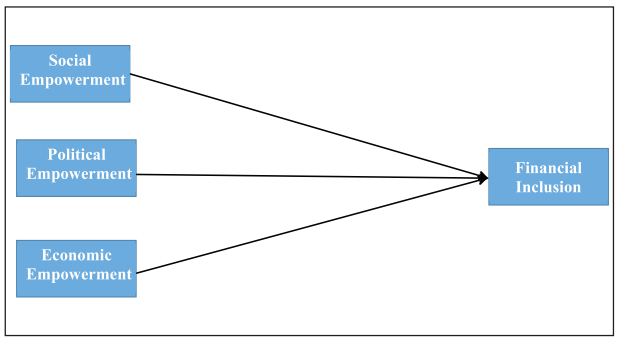


Figure 1: Framework of the Study

Thus, the particular goals of the present investigation are to explore the compass and utilization of financial inclusion schemes like PMJDY, PMJJBY, PMSBY, and APY among female slum dwellers at Ludhiana, and to examine the effects of aforementioned schemes on their social, political, and economic empowerment (as shown in Figure 1).

RESEARCH METHODOLOGY

The current study explores positive effect of financial inclusion on lives of women living in urban slums. This study is both relevant and important for conceptualizing and actualizing programmes for poor, and particularly women.

This study has actualized a qualitative plan to study reach and access of women in slums. The study was conducted between January 2018 and August 2018. There are, in total, 218 slums in Ludhiana (The location and details of slums are in Figure 2 and Table 3). As per 2011 census survey, the population of slum dwellers is 15.08 per cent of total population of the city. About three-fourths of total slums were established between the years 1981 and 2000 (www.mcludhiana.gov.in). The details of slums and their population sizes are shown in Table 3.

Table 3: Distribution of Slums by Their Population Size

S. no	Population Size of Slum	Number of Slums
1	<500	32
2	501–1,000	52
3	1,001–2,000	50
4	2,001–3,000	40
5	3,001–5,000	16
6	5,001–10,000	10

(Table 3 Continued)

(Table 3 Continued)

S. no	Population Size of Slum	Number of Slums
7	10,001–15,000	2
8	Information not available	16
		218

Source: Municipal Corporation Ludhiana.

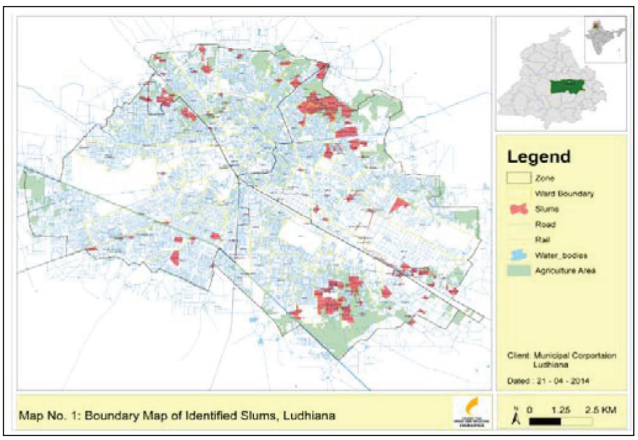


Figure 2: Map Highlighting the Slums in Ludhiana

Source: Municipal Corporation Ludhiana.

The sampling unit, that is, women in slums, was chosen using simple random sampling for this study. To test the hypothesized relationships, a questionnaire-based survey approach was used with 12 items for the constructs in this study. The questionnaire was partitioned in two sections. The first section analysed socio-economic status and access to different financial inclusion schemes of the Indian government. Second part collected the information for social, economic, and political women empowerment. The items of political empowerment were adapted from Niemi, Craig, and Mattei (1991) and Boley, Ayscue, Maruyama, and Woosnam (2016); the items of social empowerment were adapted from Basu (2006) and Boley et al. (2016); and the items of economic empowerment were adapted from Raj et al. (2018), Postmus et al. (2013) and Basu (2006). All estimations in the investigation were subjective assessments of the respondents utilizing five-point Likert scale (where 1 represents strongly disagree and 5 represents strongly agree).

The questionnaire was additionally scrutinized by two academicians and two senior financial specialists with 12 years of expertise in the area of financial inclusion. They provided remarks on the draft of the questionnaire, assessing if the items were easily understandable and carried significance.

Table 4: Profile of Respondent as Per Employment, Education, and Number of Members in Family

Employment Details		Frequency (%)
Sector	Government Sector	2
	Private Sector	69
Employment type	Full time	56
	Part Time	24
	Unemployed	20
Type of labour	Regular Labour	25
	Wage Earner	47
	Self Employed	4
	Casual Labour	27
Education profile		
1	No formal education	22
2	Education below matriculation	27
3	Matriculation/Higher Secondary	40
4	Other Technical Course	10
5	Graduation and above	1
Total family members		
1	Up to 2 members	24
2	3–5 members	64
3	6 members and above	12

Data Collection Process and Analysis

The sample consisted of women living in urban slums and the profile of respondent is provided in Table 4. Respondents for the survey were randomly selected and the eligible respondents for the survey were women owning PMJDY bank accounts. This ensured the eligibility of respondents to enrol into financial inclusion schemes. In total, 1,000 respondents were approached, of which 737 substantial responses were gathered (73% response rate). In order to identify and validate the constructs, the exploratory and confirmatory factor analyses were applied on the data. Further, to test the hypothesis, one-way analysis of variance (ANOVA) was applied.

ANALYSIS AND FINDINGS

The collected data were further analysed to know the current status of financial inclusion among female slum dwellers. Figure 3 shows the access of PMJDY accounts as per household and sources of awareness of respondents for PMJDY accounts. Figure 4 shows the frequency of operation of the PMJDY account.

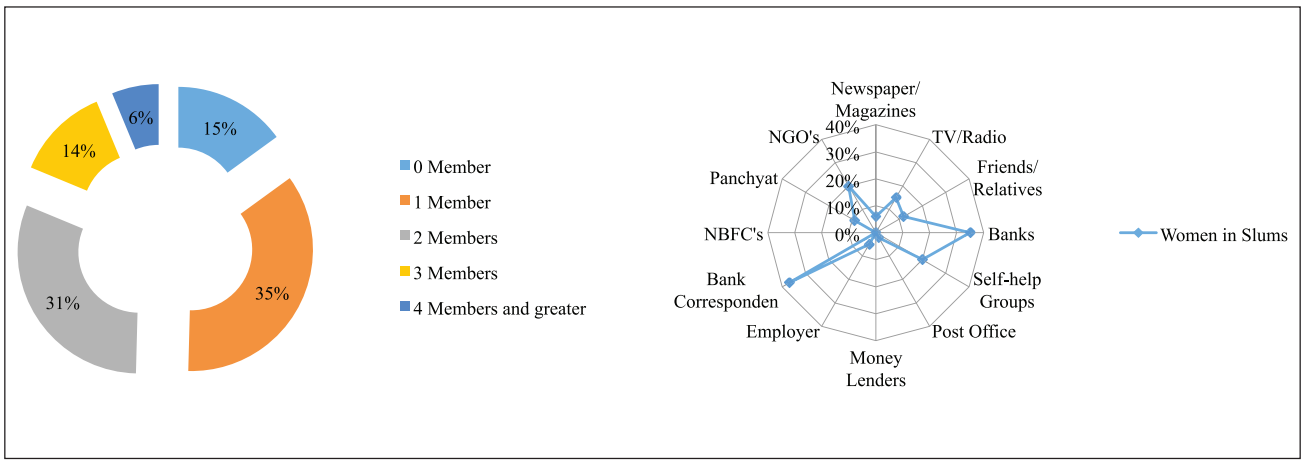


Figure 3: PMJDY Accounts as Per Households, Awareness, and Use of Accounts

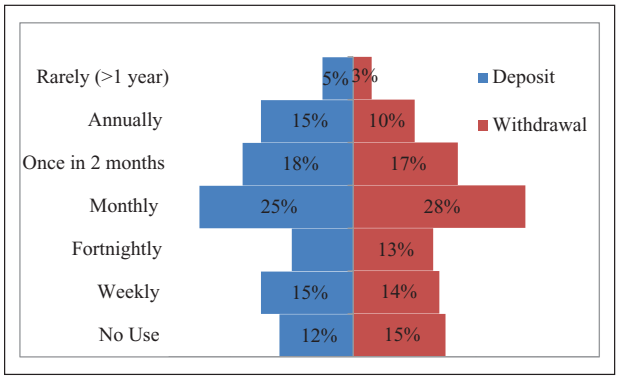


Figure 4: Frequency of Operation of PMJDY Account

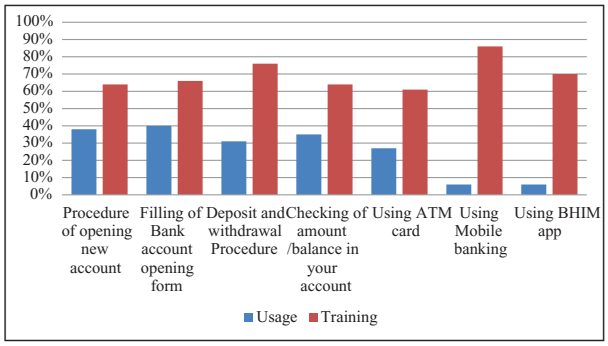


Figure 5: Usage and Training in Banking Services

Table 5: Awareness and Subscription of Insurance and Pension Schemes

Insurance, Pension, and Financial Schemes	Awareness (%)	Subscription (%)
PMSBY	40	25
PMJJBY	28	11
APY	24	5

To upgrade the level of financial inclusion in the country, the Indian government has introduced three social security schemes, namely, PMJJBY, PMSBY, and APY in the last 5 years. The goal of these schemes is to serve the objective of universal financial inclusion by penetrating insurance towards the bottom of the pyramid, that is, the weaker sections.

It was found that majority of women were aware about PMSBY and this scheme had higher subscription rate. It was also found that the general awareness and financial literacy for life insurance, accidental insurance, and pension were very low among slum dwellers at Ludhiana (Table 5).

The respondents were further intrigued regarding the usage of banking services and their training needs. It was found that 40 per cent of women at slum could fill bank account opening form, 35 per cent of women could check their account balance, and only 6 per cent of them could use mobile banking facilities. Respondents listed the need for training in mobile banking, withdrawal and deposit procedure, and operating Bharat Interface for Money (BHIM) mobile payment app (Figure 5).

Exploratory Factor Analysis (EFA)

Exploratory factor analysis was applied using principal component analysis and Varimax rotation (SPSS 22.0).

The factor analysis was suitable for this data as the Kaiser–Meyer–Olkin[1] (KMO) value was 0.861 (more than 0.7) and the Bartlett's test of Sphericity[2] was found significant at p-value<0.05. The items having factor loadings of more than 0.4 are reported in Table 5 and three items with factor loading less than 0.4 were removed from further analysis. As reported in Table 6, the factors with eigenvalue of 1.0 or more were retained for further analysis. The three extracted factors represented 72.51

per cent of the total variance explained in the data. The value of Cronbach's alpha for the three factors ranged between 0.757 (for political empowerment) and 0.922 (for economic empowerment).

Confirmatory Factor Analysis (CFA)

The multi-factor CFA was applied on 12 items to test the measurement model using the maximum likelihood method. The model fit was assessed appropriate, and the three-factor model had the best overall fit for the data with a $\chi^2/df = 1.2$ ($\chi^2 = 694.545$, df =0 564); CFI as 0.960; GFI as 0.922; AGFI as 0.901; and RMSEA as 0.078. The discriminant value was checked though average variance explained (AVE) values for all constructs were more than 0.5 and smallest items of test measurements were more than 1.96 (χ=0.05) (Fornell & Larcker, 1981). The composite reliability of each factor extended from 0.85 to 0.94 (reported in Table 5), which surpasses the suggested level of 0.6 (Hu & Bentler, 1999). The hypothesis for relationship was assessed by applying one-way ANOVA.

One-Way ANOVA

The study examines whether there is a significant mean difference in financial inclusion levels on social, political, and economic empowerment of women in slums. Financial inclusion has been divided into three levels based on accesses and usage, that is, PMJDY, PMSBY & PMJJBY, and APY. The results of one-way ANOVA showed there is significant difference between the overall financial inclusion and empowerment indicators. Games-Howell post hoc test was used to identify the difference in financial inclusion (Table 7). Table 7 shows the result of one-way ANOVA and statistically significant difference between the group means. The *p*-value for PMJDY, PMSBY, and PMJJBY for economic empowerment is 0.010, for social empowerment is 0.030, and political empowerment is 0.039, which are below 0.05. Thus, there is a statistical difference in the empowerment of the respondents who have availed PMJDY and PMSBY and PMJJY. As per the mean difference, the empowerment of the respondents has increased by availing the insurance schemes of PMSBY and PMJJY. On the other hand, the mean for economic, political, and social empowerment is not statistically significant for APY.

DISCUSSION AND CONCLUSION

This study examined the relationships between financial inclusion and women empowerment based on social, political, and economic empowerment in urban slums. The proposed hypothesis was supported by the results. Overall, the current study makes two contributions—one, advancing literature on financial inclusion of women in urban slums and two, assessing the impact of recently launched financial inclusion schemes by the Government of India.

Table 6: Rotated Component Matrix

Factor	Dimension	Statement	Factor Loading
F1	Economic empowerment	Eigenvalue = 4.084; Variance = 28.33%;AVE = 0.81; α = 0.922; CR = 0.94	
	EE1	I have say in amount of money to be saved per month	0.931
	EE2	I manage resources for repayment of credit	0.910
	EE3	I work to generate further income	0.914
	EE4	I participate in decision about the usage of savings and credit use	0.844
F2	Social empowerment	Eigenvalue = 2.665; Variance = 24.23%; AVE = 0.71; α = 0.863; CR = 0.90	
	SE1	Important decisions in the family are taken mutually by me and my husband	0.830
	SE2	I can move out of home freely without any restrictions for work	0.841
	SE3	Working freely has improved my social status.	0.908
	SE4	I know most of the women living in my locality	0.790
F3	Political empowerment	Eigenvalue = 1.953; Variance = 19.57%; AVE = 0.60; α = 0.757; CR = 0.85	
	PE1	I have knowledge of parties operating in the area	0.757
	PE2	I have good understanding of important political issues of country	0.861
	PE3	My vote make difference to development in my area	0.722
	PE4	I participate in political discussion in family	0.727

Notes: Total variance = 72.51%; KMO = 0.861; chi-square = 323.988; Df = 66; Sig = 0.00.

F1 = Economic Empowerment; F2 = Social Empowerment; F3 = Political Empowerment.

Table 7: Results of One-Way ANOVA and Post-hoc Test

Characteristics	Factor		Categories				
			1	2	3	*F*-Value	*p*-Value
Economic empowerment	Financial inclusion	N	426	273	38	62.262	0.018
		Mean	13.04	14.5	16.		
		SD	3.4	2.8	3.2		
Post- hoc test							
Financial inclusion (i)	Financial inclusion (J)	Mean difference (i–j)					
2	1	1.53837					0.010
	3	−1.98423					0.330
Social empowerment	Financial inclusion	N	426	273	38	41.211	0. 028
		Mean	12.9	14.1	14.9		
		SD	2.6	2.2	1.9		
Post-hoc test							
Financial inclusion (i)	Financial inclusion (J)	Mean difference (i–j)					
2	1	1.21096					0.030
	3	−0.82889					0.200
Political empowerment	Financial inclusion	N	426	273	38	69.175	0.039
		Mean	12.9	13.5	15.4		
		SD	2.4	2.6	1.9		
Post-hoc test							
Financial Inclusion (i)	Financial Inclusion (J)	Mean Difference (i–j)					
2	1	1.40300					0.009
	3	−1.43174					0.150

As recognized in this study, all sorts of financial inclusion processes have the ability to impact the empowerment of women and this is in line with the available literature which suggests that financial inclusion has an impact on the social empowerment of women (Al-Mamun et al., 2014; Basu, 2006; Blattman, Green, Annan, & Jamison, 2013; Chatterjee et al., 2018; Goetz & Gupta, 1996; Rahaman, 1999; Weber & Ahmad, 2014; Wrigley-Asante, 2012). Similarly, the results of financial inclusion and political empowerment (Basu, 2006; Bhattacharyya, 2019; Hashemi, Schuler, & Riley, 1996; Kabeer, 1999; Malhotra & Schuler, 2005) and economic empowerment (Datta & Sahu, 2017; Bayulgen, 2015; Hashemi et al., 1996; Kabeer, 1999) are consistent with the literature. Consequently, this study presents an all-encompassing replication of well-built inclusive-credibility connections. Replication studies with extensions are important to break down the use of results on a wide scale (Hermes & Lensink, 2011) to identify generalization of relationships, for both with Indian population of slum dwellers and in context of financial inclusion.

The outcomes reveal that distinctive components of empowerment are statistically significant by joining formal financial services. Women empowerment is related to each one of the measures of empowerment which reverberate the hypotheses in this study and past investigations. The study also reveals that women with higher access and usage of financial services, such as opening bank account and availing insurance, have higher social, political, and economic empowerment. This has implications for increasing women's bargaining power in society, that is, greater negotiations, freedom for political choices, taking financial decisions at work, and decision-making within family, leading to higher women empowerment.

The results of the current investigations regarding the effect of financial inclusion on empowerment dimensions for women in urban slums also indicate that women with access to PMSBY and PMJJBY have higher overall social, political, and economic empowerment. However, women accessing APY show no significant impact on the social, political, and economic empowerment. One of the plausible reasons

may be that women with higher pool of savings have keen interest in inclusive insurance (Fletschner & Kenney, 2014). It may be due to the financial coverage, especially in the time of disaster and risk.

The results suggest that the PMJDY scheme, launched in 2014–2015 by the Government of India as a part of financial inclusion strategy, has been quite successful especially in case of women in slums. It was noteworthy that even less educated women have joined formal banking system and had sound awareness regarding these financial inclusion schemes. However, the subscriptions to these schemes were not considerable. Both, non-profit organizations and SHGs, play a vital role in increasing awareness regarding these schemes among women in slums. However, the digital payments and usage of mobile banking applications are still a challenge. The Indian government should create a supportive environment to enhance the impact of these schemes.

The second objective was to confirm the convergence of dimensions of empowerment for women in urban slums. The strong relationships between women empowerment and financial inclusion provide insights to policy makers.

Third objective was to show the linkage between financial inclusion and dimension of women empowerment. A positive influence of financial inclusion on social, political, and economic dimensions of women empowerment was found. But the long-term sustainability of these schemes relies upon its involvement in generation of economic activities for women. The policy decisions should accelerate procedures through supportive mechanism, and develop superior systems and easy interaction on platforms among different stakeholders.

This study also revealed that women in urban slums did not lack access to financial institutions but suffered from (a) economic risks such as unsteady earnings, non-contractual informal jobs, (b) financial risks such as reliance on cash economy, informal credit, and (c) social risks such as social fragmentations. Considering these differences, there are substantial gaps in the lives of poor urban women. In order to bridge this gap, there is a need to systemically understand financial lives of these households and customize the delivery mode such that it is affordable and scalable. Thus, there is a need to create new points for information collection, so that more inclusive financial system for women in slums may be created.

This study makes several contributions to theory and practice. The study advances the debate on women empowerment of women ghettoes through financial inclusion schemes of the Government of India. The study identifies the substantial need for the development of formal financial system that may enhance the scale of financial inclusion.

This study has certain limitations. Data were collected from the industrial city of Ludhiana (Punjab) that restricts the generalization of results to the Indian population. The questionnaire approach was the only way to gather primary data and thus, the results might have a common-method bias. This research does not consider longitudinal research structure and is cross-sectional in nature which, in turn, does not permit considering long haul interactions between women empowerment and financial inclusion measures.

A pan-India, city and state study to examine whether comparable outcomes can be produced. This will provide an all-encompassing perspective useful in planning future research strategy at national level. An exhaustive qualitative research in the future may create a wide-going thought regarding financial inclusion and women empowerment.

DECLARATION OF CONFLICTING INTERESTS

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

The authors received no financial support for the research, authorship, and/or publication of this article.

NOTES

1. The Kaiser–Meyer–Olkin measure of sampling adequacy is a statistic that indicates the proportion of variance in your variables that might be caused by underlying factors. High values (close to 1.0) generally indicate that a factor analysis may be useful with your data. If the value is less than 0.50, the results of the factor analysis probably will not be very useful.
2. Bartlett's test of sphericity tests the hypothesis that your correlation matrix is an identity matrix, which would indicate that your variables are unrelated and therefore unsuitable for structure detection. Small values (<0.05) of the significance level indicate that a factor analysis may be useful with your data.

ORCID ID

Shivangi Bhatia https://orcid.org/0000-0001-5866-0405

REFERENCES

Afrin, S. (2008). A multivariate model of micro credit and rural women entrepreneurship development in Bangladesh. *International Journal of Business and Management, 16(1), 169–185.*

Agier, I., & Szafarz, A. (2013). Subjectivity in credit allocation to micro-entrepreneurs: Evidence from Brazil. *Small Business Economics, 41*(1), 263–275.

Ahmed, J., Aurora, A., Biru, A., & Salvini, S. (2001). *Gender issues, population and development in Ethiopia. Addis Ababa, Ethiopia: CSA.*

Al-Mamun, A., Wahab, S. A., Mazumder, M. N. H., & Su, Z. (2014). Empirical investigation on the impact of microcredit on women empowerment in urban Peninsular Malaysia. *The Journal of Developing Areas, 48*(2), 287–306.

Ambarkhane, D., Singh, A. S., & Venkataramani, B. (2014). Measuring financial inclusion of Indian states. *International Journal of Rural Management, 12(1), 72–100.*

Armendáriz, B., & Morduch, J. (2010). *The economics of microfinance*. MIT press.

Arora, R. U. (2010). *Measuring financial access. Discussion Paper Economics, No-7, Griffith University.*

Ashraf, N., Karlan, D., & Yin, W. (2010). Female empowerment: Further evidence from a commitment savings product in the Philippines. *World Development, 28*(3), 333–344.

Bagil, S., & Dutta, P. (2012). A study of Financial Inclusion in India. *Journal of Radix International Educational and Research Consortium, 8*(1), 1–18.

Bandiera, O., Buehren, N., Burgess, R., Goldstein, M., Gulesci, S., Rasul, I., & Sulaiman, M. (2018). *Women's empowerment in action: Evidence from a randomized control trial in Africa*. Washington, DC: World Bank.

Bapat, D., & Bhattacharyaya, B. N. (2016). Determinants of financial inclusion of urban poor in India: *An empiricalÚnalysis. CESIFO Working Paper, 6096*, 1-31.

Bardhan, K., & Klasen, S. (1999). UNDP's gender-related indices: A critical review. *World Development, 27*(6), 985–1010.

Barik, R., & Sharma, P. (2019). Analyzing the progress and prospects of financial inclusion in India. *Journal of Public Affairs, e1948*, 1-6.

Basu, J. P. (2006). *Microfinance and women empowerment: An empirical study with special reference to West Bengal*. Mumbai, India: Indira Gandhi Institute of Development Research.

Basu, P., & Srivastava, P. (2005). Exploring possibilities: Microfinance and rural credit access for the poor in India. *Economic and Political Weekly, 40*(17), 1747–1756.

Bateman, M. (2012). Let's not kid ourselves that financial inclusion will help the poor. *The Guardian*, p. 8.

Batliwala, S. (2007). Taking the power out of empowerment—An experiential account. *Development in Practice, 17*(4–5), 557–565.

Bayulgen, O. (2015). Microcredit and political empowerment in Azerbaijan and Kazakhstan. *International Journal of Development Issues, 14*(2), 130–148.

Beteta, H. C. (2006). What is missing in measures of women's empowerment? *Journal of Human Development, 7*(2), 221–241.

Bhanot, D., Bapat, V., & Bera, S. (2012). Studying financial inclusion in North-East India. *International Journal of Bank Marketing, 30*(6), 465–484.

Bhattacharyya, R. (2019). Microfinance: A tool for empowering women in India. In R. C. Das (Ed.), *Handbook of research on microfinancial impacts on women empowerment, poverty, and inequality* (pp. 23–51). Hershey, PA: IGI Global.

Bhutoria, A., & Vignoles, A. (2018). Do financial education interventions for women from poor households impact their financial behaviors? Experimental Evidence from India. *Journal of Research on Educational Effectiveness, 11*(3), 409–432.

Bihari, S. C. (2011). Growth through financial inclusion in India. *Journal of International Business Ethics, 5(1), 28–41.*

Blattman, C., Green, E., Annan, J., & Jamison, J. (2013). *Building women's economic and social empowerment through enterprise: An experimental assessment of the Women's income generating support program in Uganda*. Washington, DC: The World Bank.

Boley, B. B., Ayscue, E., Maruyama, N., & Woosnam, K. M. (2016). Gender and empowerment: Assessing discrepancies using the resident empowerment through tourism scale. *Journal of Sustainable Tourism, 25*(1), 113–129.

Brana, S. (2012). Microcredit: An answer to the gender problem in funding? *Small Business Economics, 40*(1), 87–100.

Bylander, M. (2014). Borrowing across borders: Migration and microcredit in rural Cambodia. *Development and Change, 45*(2), 284–307.

Chaia, A., Dalal, A., Goland, T., Gonzalez, M. J., Morduch, J., & Schiff, S. (2009). Half the world is unbanked: State of financial inclusion policy in developing countries AFI. *Financial Access Initiative, Framing Note*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.401.5164&rep=rep1&type=pdf

Chakravarty, S. R., & Pal, R. (2013). Financial inclusion in India: An axiomatic approach. *Journal of Policy Modeling, 35*(5), 813–837.

Chatterjee, S., Dutta Gupta, S., & Upadhyay, P. (2018). Empowering women and stimulating development at bottom of pyramid through micro-entrepreneurship. *Management Decision, 56*(1), 160–174.

Chiapa, C., Prina, S., & Parker, A. (2016). The effects of financial inclusion on children's schooling, and parental aspirations and expectations. *Journal of International Development, 28*(5), 683–696.

COHRE. (2008). *Women, slums and urbanisation: Examining the causes and consequences. Retrieved from* http://globalinitiative-escr.org/wp-content/uploads/2013/05/women_slums_and_urbanisation_may_2008.pdf

Corsi, M., & De Angelis, M. (2017). Gender discrimination in microfinance? Some evidence from Uganda. *The Journal of Development Studies, 53(5), 723–740.* doi:10.1080/00220388.2016.1205733.

Datta, S. K., & Singh, K. (2018). Aspects of inclusion and peoples' empowerment related to national rural employment Guarantee Scheme in India. *International Journal of Public Administration, 41*(2), 83–94.

Datta, S., & Sahu, T. N. (2017). An empirical study on the impact of microfinance on women empowerment: Evidence from West Bengal. *Indian Journal of Commerce and Management Studies, 8*(3), 53.

Dawar, G., & Singh, S. (2016). Corporate social responsibility and gender diversity: A literature review. *Journal of IMS Group, 13(1), 61–71.*

Deininger, K., & Liu, Y. (2013, March) Economic and social impacts of an innovative self-help group model in India. *World Development, 43*, 149–163.

Demirguc Kunt, A., Klapper, L., & Singer, D. (2013). *The Global Findex database: Financial inclusion in India Findex notes, no. 8. Washington DC: World Bank.*

Demirgüç-Kunt, A., Klapper, L. (2012). *Measuring financial inclusion. The Global Findex database* (Policy Research Working Paper No. 6025). Washington, DC: The World Bank.

Dev, S. M. (2006). Financial inclusion: Issues and challenges. *Economic and Political Weekly, 41*(41), 4310–4313.

Duvendack, M., Palmer-Jones, R., Copestake, J. G., Hooper, L., Loke, Y., & Rao, N. (2011). *What is the evidence of the impact of microfinance on the well-being of poor people? London, UK: EPPI-Centre, University of London.*

Dymski, G. A. (2005). Financial globalization, social exclusion and financial crisis. *International Review of Applied Economics, 19*(4), 439–457.

Fernandes, D., Lynch, J. G. Jr., & Netemeyer, R. G. (2014) Financial literacy, financial education, and downstream financial behaviors. *Management Science, 60*(8), 1861–1883.

Fisher, J. (2008). Women in water supply, sanitation and hygiene programmes. Proceedings of the ICE: *Municipal Engineer, 161*(4), 223-229

Fletschner, D. (2008). Women's access to credit: Does it matter for household efficiency? *American Journal of Agricultural Economics, 90*(3), 669–683.

Fletschner, D. (2009). Rural women's access to credit: Market imperfections and intra-household dynamics. *World Development, 37*(3), 618–631.

Fletschner, D., & Kenney, L. (2014). *Rural women's access to financial services: Credit, savings, and insurance, in gender in agriculture* (pp. 187–208). Dordrecht, The Netherlands: Springer.

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research, 18*(3), 382–388.

Fraser, N. (2011). Feminism, capitalism, and the cunning of history. *New Left Review, 56*, 97–117.

Frisancho, V. (2016). How to raise household savings in LAC: Constraints and best practices. IDB Technical Note (Department of Research and Chief Economist) IDB-TN-962. Washington, DC: Inter-American Development Bank.

Fuller, D. (1998). Credit union development: Financial inclusion and exclusion. *Geoforum, 29*(2), 145–157.

Ghosh, S., & Vinod, D. (2017, April). What constrains financial inclusion for women? Evidence from Indian micro data. *World Development, 92*, 60–81.

Goetz, A. M., & Gupta, R. S. (1996). Who takes the credit? Gender, power, and control over loan use in rural credit programs in Bangladesh. *World Development, 24*(1), 45–63.

Guérin, I., Kumar, S., & Agier, I. (2013). Women's empowerment: Power to act or power over other women? Lessons from Indian microfinance. *Oxford Development Studies, 41*(sup1), S76–94.

Guérin, I., Labie, M., & Servet, J. M. (2015). *The crises of microcredit.* Chicago, IL: University of Chicago Press.

Hashemi, S. M., Schuler, S. R., & Riley, A. P. (1996). Rural credit programs and women's empowerment in Bangladesh. *World Development, 24*(4), 635–653.

Hazarika, G., & Guha-Khasnobis, B. (2008). *Household access to microcredit and children's food security in rural Malawi: A gender perspective* (Discussion Paper Series IZA DP No 3793). Bonn, Germany: Institute for the Study of Labor (IZA).

Hazarika, I. (2010). Women's reproductive health in slum populations in India: Evidence from NFHS-3. *Journal of Urban Health, 87*(2), 264–277.

Hermes, N., & Lensink, R. (2011). Microfinance: Its impact, outreach, and sustainability. *World Development, 39*(6), 875–881.

Holloway, K., Niazi, Z., & Rouse, R. (2017). *IPA women's economic empowerment through financial inclusion: A review of existing evidence and remaining knowledge gaps.* New Haven, CT: Innovations for Poverty Action.

Hu, L., & Bentler, P. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling, 6*(1), 1–55.

Hummel, A. (2013). The Commercialization of microcredits and local consumerism: Examples of over-indebtedness from indigenous Mexico. In I. Guérin, S. Morvant-Roux, & M. Villarreal (Eds.), *Microfinance, debt and over-indebtedness: Juggling with money (pp. 253–271).* London, UK: Routledge.

Kabeer, N. (1999). Resources, agency, achievements: Reflections on the measurement of women's empowerment. *Development and Change, 30*(3), 435–464.

Karim, L. (2011). *Microfinance and its discontents: Women in debt in Bangladesh. Minneapolis, MN: U*niversity of Minnesota Press.

Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management Science, 62(12), 3393–3411.*

Karlan, D., Ratan, A. L., & Zinman, J. (2014). Savings by and for the Poor: A research review and agenda. *The Review of Income and Wealth, 60*(1), 36–78.

Kaur, M., Singh, S., Gupta, M., Bahuguna, P., & Rani, S. (2015). Inequity in access to health services between migrants and natives of Chandigarh, India. *International Journal of Migration, Health and Social Care, 11*(2), 147–155.

Kim, J. H. (2016). A study on the effect of financial inclusion on the relationship between income inequality and economic growth. *Emerging Markets Finance and Trade, 52*(2), 498–512.

Klasen, S. (2002). Low schooling for girls, slower growth for all? *Cross*‐country evidence on the effect of gender inequality in education on economic development. *The World Bank Economic Review, 16*(3), 345–373.

Klasen, S., & Lamanna, F. (2009). The impact of gender inequality in education and employment on economic growth: New evidence for a panel of Countries. *Feminist Economics, 15*(3), 91–132.

Kumar, C., & Mishra, S. (2009). *Banking outreach and household level access: Analyzing financial inclusion in India.* Mumbai, India: IGIDR.

Kumar, N., Raghunathan, K., Arrieta, A., Jilani, A., Chakrabarti, S., Menon, P., & Quisumbing, A. R. (2019). Social networks, mobility, and political participation: The potential for women's self-help groups to improve access and use of public entitlement schemes in India. *World Development, 114*, 28–41.

Laha, A. (2015). Association between financial inclusion and human development in South Asia: A cross-country analysis with special reference to India. *Journal of Economic Policy and Research, 10*(2), 69–91.

Laha, A., & Kuri, P. K. (2014). Demand for and supply of financial inclusion in India: An inter-state analysis. *ELK Asia Pacific Journal of Finance and Risk Management, 5*(1), 103–116.

Lampietti, J. A., & Stalker, L. (2000). *Consumption expenditure and female poverty a review of the evidence, policy research report on gender and development* (Working Paper Series No. 11). Washington, DC: The World Bank.

Lenka, S. K., & Bairwa, A. K. (2016). Does financial inclusion affect monetary policy in SAARC countries? *Cogent Economics & Finance, 4*(1), 1127011.

Lenka, S. K., & Barik, R. (2018). A discourse analysis of financial inclusion: Post-liberalization mapping in rural and urban India. *Journal of Financial Economic Policy, 10*(3), 406–425.

Lenka, S. K., & Sharma, R. (2017). Does financial inclusion spur economic growth in India? *The Journal of Developing Areas, 51*(3), 215–228.

Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature, 52*(1), 5–44.

Malhotra, A., & Schuler, S. R. (2005). Women's empowerment as a variable in international development. In D. Narayan-Parker (Ed.), *Measuring empowerment: Cross-disciplinary perspectives* (pp. 71–88). Washington, DC: The World Bank.

Mason, K. O. (2005). Measuring women's empowerment: Learning from cross-national research. In D. Narayan-Parker (Ed.), *Measuring empowerment: Cross-disciplinary perspectives* (pp. 89–102). Washington, DC: The World Bank

Mehrotra, A., & Nadhanael, G. V. (2016). Financial Inclusion and Monetary Policy in Emerging Asia. *In Financial Inclusion in Asia* (pp. 93-127). Palgrave Macmillan, London.

Muravyev, A., Talavera, O., & Schäfer, D. (2009). Entrepreneurs' gender and financial constraints: Evidence from international data. *Journal of Comparative Economics, 37*(2), 270–286.

Nasrin, F. (2012). Women, environment and environmental advocacy: Challenges for Bangladesh. *Asian Journal of Social Sciences and Humanities, 1*(3), 1–3.

Niemi, R. G., Craig, S. C., & Mattei, F. (1991). Measuring political efficacy in the 1988 national election study. *American Political Science Review, 85*(4), 1407–1413.

OECD. (2011). *Women's economic empowerment.* Retrieved October 6, 2019, from https://www.oecd.org/dac/gender-development/47561694.pdf

Page, N., & Czuba, C. E. (1999). Empowerment: What is it. *Journal of Extension, 37*(5), 1–5.

Park, C. Y., & Mercado, R. V. (2016). Does financial inclusion reduce poverty and income inequality in developing Asia? In S. Gopalan & T. Kikuchi (Eds.), *Financial inclusion in Asia* (pp. 61–92). London, UK: Palgrave Macmillan.

Peachy, S., & Roe, A. (2006). *Access to finance: What does it mean and how do savings banks foster access*. Bruxelas, Belgium: World Savings Banks Institute (WSBI).

Pillai, T. J. (2016). Financial inclusion and Pradhan Mantri Jan Dhan Yojana: An empirical study conducted in the villages of Mulshi Taluka in Pune district. *Indian Journal of Science and Technology, 9*(45), 1–6.

Postmus, J. L. (2010, October). *Economic empowerment of domestic violence survivors*. Harrisburg, PA: VAWnet Applied Research Forum, National Resource Center on Domestic Violence.

Postmus, J. L., Plummer, S. B., McMahon, S., & Zurlo, K. A. (2013). Financial literacy: Building economic empowerment with survivors of violence. *Journal of Family and Economic Issues,* 34(3), 275–284.

Quisumbing, A. R., Haddad, L., & Pena, C. (1995). *Gender and poverty: New evidence from 10 developing countries* (FCND Discussion Paper No. 9). Washington, DC: International Food Policy Research Institute.

Rahman, A. (1999). Micro-credit initiatives for equitable and sustainable development: who pays?. *World development, 27*(1), 67-82.

Raichoudhury, A. (2016). Financial inclusion & human development: A cross-country analysis. *Asian Journal of Business Research, 6*(1), 34–48.

Raj, A., Silverman, J. G., Klugman, J., Saggurti, N., Donta, B., & Shakya, H. B. (2018, January). Longitudinal analysis of the impact of economic empowerment on risk for intimate partner violence among married women in Rural Maharashtra, India. *Social Science & Medicine, 196*, 197–203.

Rangarajan Committee. (2008). *Report of the committee on financial inclusion.* Government of India.

Rawlings, L. B., & Rubio, G. M. (2005). Evaluating the impact of conditional cash transfer programs. *World Bank Researcher Observer, 20*(1), 29–55.

Roodman, D., & Qureshi, U. (2006). *Microfinance as business.* Washington, *DC:* Center for Global Development.

Sen, A. (2014). Development as freedom (1999). In J. T. Roberts, A. B. Hite, & N. Chorev (Eds.), *The globalization and development reader: Perspectives on development and global change* (2nd ed., pp. 525–547). Wiley.

Shah, M., Rao, R., & Shankar, P. S. V. (2007). Rural credit in 20th century India: Overview of history and perspectives. *Economic & Political Weekly, 42*(15), 1351–1365.

Siddiqui, T., & Siddiqui, K. I. (2017). Determining effect of telecom services on financial inclusion in Haryana using SEM. In T. A. Siddiqui & K. I. Siddiqui (Eds.), *Proceedings of international conference on research and business sustainability* (pp. 436–441).

Singh, C., & Naik, G. (2018). Financial inclusion after PMJDY: A case study of Gubbi Taluk, Tumkur (IIM Bangalore Research Paper No. 568). Retrieved from https://ssrn.com/abstract=3151257 or http://dx.doi.org/10.2139/ssrn.3151257

Sinha, F. (2006). Social rating and social performance reporting in microfinance. *Towards a common framework, the SEEP network for the Argidius Foundation*. Retrieved from http://www.m-cril.com/wp-content/uploads/2017/06/Social-Rating-and-Social-Performance-Reporting-in-Microfinance-2006.pdf

Stewart, R., Van Rooyen, C., Korth, M., Chereni, A., Da Silva, N. R., & De Wet, T. (2012). *Do micro-credit, micro-savings and micro-leasing serve as effective financial inclusion interventions enabling poor people, and especially women, to engage in meaningful economic opportunities in low-and middle-income countries? A systematic review of the evidence. EPPI-Centre.*

Sundström, A., Paxton, P., Wang, Y. T., & Lindberg, S. I. (2017, June). Women's political empowerment: A new global index, 1900–2012. *World Development, 94*, 321–335.

Swain, R. B., & Varghese, A. (2009). Does self-help group participation lead to asset creation? *World Development, 37*(10), 1674–1682.

Swain, R. B., & Wallentin, F. Y. (2009). Does microfinance empower women? Evidence from Self Help Groups in India. *International Review of Applied Economics, 23*(5), 541–556.

Swamy, V. (2014, April). Financial inclusion, gender dimension, and economic impact on poor households. *World Development, 56,* 1–15.

Taylor, J. (2000). *So now they are going to measure empowerment!* Cape Town, South Africa: Community Development Resource Association. Retrieved from www.cdra.org.za/articles.htm; www.cdra.org.za/ articles/So%20now%20they%20are%20going%20to%20measure%20empowerment!%20-% 20by%20James%20Taylor.htm

The World Bank. (2018). *Financial inclusion is a key enabler to reducing poverty and boosting prosperity.* Retrieved from https://www.worldbank.org/en/topic/financialinclusion/overview

Tulasi, G., Golait, R., Sethi, M., & Goel, S. (2017). Involuntary exclusion and the formal financial sector. *Economic and Political Weekly, 52(36), 67–72.*

United Nations. (2018). *Women's political empowerment and leadership.* Retrieved October 6, 2019, from https://www2.unwomen.org/-/media/field office eseasia/docs/publications/2018/11/fiji-mco_wpel_oct2018.pdf?la=en&vs=55

Verma, Y., & Garg, P. (2016). Pradhan Mantri Jan Dhan Yojana (PMJDY): A step towards eradicating financial untouchability. *Indian Journal of Finance, 10*(1), 56–64.

Weber, O., & Ahmad, A. (2014, October). Empowerment through microfinance: The relation between loan cycle and level of empowerment. *World Development, 62*, 75–87.

Wrigley-Asante, C. (2012). Out of the dark but not out of the cage: Women's empowerment and gender relations in the Dangme west district of Ghana. *Gender, Place & Culture, 19*(3), 344–363.

Yunus, M. (2004). Grameen bank, microcredit and millennium development goals. *Economic and Political Weekly, 39*(36), 4077–4080.

Shivangi Bhatia is a research scholar at Delhi Technological University. She also teaches financial accounting, laws in banking, decision science, and financial statement analysis. Her areas of interest are gender studies and financial inclusion.

e-mail: shivangibhatia88@gmail.com

Seema Singh is professor of economics at the Delhi Technological University. She is interested in gender, higher education, employment, and labour market. She has presented and published several papers. She has also been involved in several research projects. She received AICTE Career Award for Young Teachers in 2000. She has also been elected Vice President (Education and Research) and Board Member (South Asia) of International Network of Women Engineers and Scientists (INWES) for 2018–2020. She is member of the editorial team of many journals.

e-mail: seemahumanitiesdtu@gmail.com

Int J Syst Assur Eng Manag
https://doi.org/10.1007/s13198-019-00935-1

ORIGINAL ARTICLE

Ensemble-based extreme learning machine model for occupancy detection with ambient attributes

Sachin Kumar[1] · **Jagvinder Singh**[2] · **Ompal Singh**[3]

Received: 2 September 2019 / Revised: 2 September 2019
© The Society for Reliability Engineering, Quality and Operations Management (SREQOM), India and The Division of Operation and Maintenance, Lulea University of Technology, Sweden 2020

Abstract Context-aware computing is a growing research domain in present circumstances due to technological advancements in the area of sensors technology, big data, artificial intelligence and robotics and automation. It has many applications for making the daily life of human beings sustainable, comfortable, and smooth. Context ware computing also includes ambient intelligence and applications such as occupancy detection, prediction, user recognition etc. Occupancy detection and recognition help in developing intelligent applications which help the energy management, intelligent decision making, that results in cost reduction and fault and failure prevention of services and products in advance. Several studies have been conducted to detect the occupancy with a different set of methodologies and approaches using varying types of data such as environmental parameters, image and video-based attributes, wireless or sensor based parameters, and noise-based parameters. This paper proposes a reliable, more accurate and efficient model based on the statistical analysis of the sensor based data for occupancy detection. Detailed quantification of the relationship of the ambient attributes is presented and the ensemble model is developed based on machine learning technique extreme learning machine to achieve the significant level of improvement in accuracy, efficiency, generalization and reliability. In addition to this, the paper also proposes one online and adaptive model-based online sequential extreme learning machine to perform occupancy detection on real-time data when complete data is not available and learning is done with recent data points coming in the form of streams. Results are compared with existing work in the domain and it is observed that proposed model perform better in terms of efficiency and accuracy over existing literature work.

Keywords Classification · ELM · Ensemble based model · Occupancy detection · OSELM

1 Introduction

Detection of the occupancy of the building's space used of residence or commercial space is fundamental to many applications such as home automation, energy consumption and electricity consumption optimization, surveillance and security applications, human behaviour analysis, dynamic space allocation and resources utilization optimization. An accurate and efficient occupancy detection system can help optimisation of usage of resources in the building using ambient intelligence obtained from processing and analysing the data coming from several sources. Additionally, Occupancy detection can also help in the development of the personalised environment as per the user's behaviour and requirements. According to several studies, the built environment is at present a major source of energy,

✉ Sachin Kumar
skumar@cic.du.ac.in; sachin.blessed@gmail.com

Jagvinder Singh
jagvinder.singh@gmail.com

Ompal Singh
drompalsingh1@gmail.com

[1] Cluster Innovation Centre, University of Delhi, Delhi, India

[2] Department of Management, Delhi Technological University, Delhi, India

[3] Department of Operation Research, University of Delhi, Delhi, India

Published online: 10 January 2020

accounting for more than 30% of the total final energy consumed in countries (Bhattacharya et al. 2016; Pérez-Lombard et al. 2008; Zheng et al. 2014). Occupancy is one of the important factors in building environment energy consumption and many studies reported that, the prediction of occupancy status in buildings can help in estimation and in saving of energy from 30 to 42% (Dong 2009; Erickson et al. 2014, 2011). Experiments conducted by Brooks et al. (2014) has shown that energy saving was about 37% and varying from 29 to 80% (Brooks et al. 2015) when occupancy related insights were used as an input for heating ventilation air conditioning(HVAC) control system and algorithms. We can take many decisions on the basis of the occupancy status of the buildings if it is accurately measured to a satisfaction level of accuracy. Applications based on such measures are called occupant driven demand control measures. Some of them are the control of thermal, visual and indoor air quality providing systems and these systems account for over 30% of the energy consumption of the office building. This demand controls measures and their association with occupancy detection can be incorporated and it will help in demand driven measures performance to improve in the ambient environment. Nowadays sensors are affordable and this has made energy-related sensors approach quiet ubiquitous and simple which combined with sufficiently affordable computing for automation systems help in innovative research findings in this research domain (Lamnabhi-Lagarrigue et al. 2017). There are many methodologies to detect occupancy and status such as image and video-based detection systems, electricity based detection systems, several types of sensors such as environmental sensors based detection, which includes, CO_2 detection sensors, PIR sensors, ultrasonic sensors, image sensors, sound sensors (Labeodan et al. 2015). Sensor-based methodology to detect accurately detect the occupancy status and the number of occupants is considered a good approach as compared to image and video processing methodology (Wang et al. 2018). The reason is that it does not create privacy and security related issues. In this paper, the data used is basically environment factors recorded with the help of the sensors. These environment-related factors are light intensity, temperature values, humidity values and CO_2 quantification. Simultaneously for occupancy, a digital camera was used to know about the occupancy on the ground. This resulted in the binary data set occupied or not occupied. Additionally, sensors based approach is cost effective and does not require high investment and infrastructure. In such an environment, Occupancy detection systems face some challenges of reliability, adaptability, efficiency and accuracy. In this direction, models based on emerging machine learning approaches might be helpful. This paper proposes novel models based on machine learning algorithm extreme learning machine(ELM) and its variants which works fine for static and dynamic data in an adaptive and real environment. ELM is an artificial neural network(ANNs) based machine learning technique which is based on the concept of single feedforward neural networks(SFNNs). ELM has been used in classification and regression applications such as temperature forecasting, energy forecasting of buildings and in many more applications (Ding et al. 2015). This work also deals with real-time occupancy detection based on the online sequential extreme learning machine(OSELM) a variant of ELM and compares results with the previous work. The proposed ensemble model improves the accuracy and efficiency over the previous research work.

Rest of the paper is organised as follows. Section 2 deals with recent work and literature survey. Section 3 describes the foundation of the modelling approaches. Section 4 discusses the results and Sect. 5 present the conclusion and future directions.

2 Literature survey

Several research studies and experiments have been conducted from a different perspective in occupancy domain (Nguyen and Aiello 2013; Kéry et al. 2010; Yang et al. 2016). Richardson et al. (2008) developed a stochastic occupancy detection model based on the survey methodology. Different from this, a model to detect occupancy was developed from digital video cameras, passive infrared detection and CO_2 sensors by Meyn et al. (2009). This proposed model used Bayesian statistics and accounted for the role of the previous information in occupancy detection. The researchers reported that their model has reduced the average error from 70 to 11%. Two models were developed to predict occupancy by Erickson et al. (2009) in which the first model was based on the multivariate Gaussian distribution(MGD) to the data obtained from a digital camera and the second model is based on an agent-based model (ABM) that can be used for simulating mobility patterns. Further, some researchers tried to detect the number of occupants in the room and model was developed using data for CO_2, temperature and ventilation to result in dynamic model (Ebadat et al. 2013). This developed model showed the occupancy level of 88% and was better than predictions obtained with support vector machine(SVM) and neural network estimators based models. Recently, ELM bases models have been used in many context-aware applications and such models are accurate and efficient and adaptive to real-time needs. In such settings, Kumar et al. (2018d) presented a conceptual for health systems based on sensor and human-driven information. Kumar et al. presented a model to forecast the

Table 1 Literature survey and other details

Author's name	Year	Accuracy	References
Robert H. Dodiera et al.	2006	76	Dodier et al. (2006)
R. N. Tomastik et al.	2008	94	Tomastik et al. (2010)
Richardson et al.	2008	87	Richardson et al. (2008)
Khee Poh Lam et al.	2009	77.6	Zhang et al. (2012)
Bing Dong et al.	2010	90	Dong et al. (2010)
Zheng Yang et al.	2012	77.04	Yang et al. (2012)
Tobore Ekwevugbe et al.	2013	84.85	Beltran et al. (2013)
Alex Beltran et al.	2013	88.5	Beltran et al. (2013)
Wilhelm Kleiminger et al.	2013	88	Kleiminger et al. (2013)
Zheng Yang et al.	2013	94	Yang et al. (2014)
Afrooz Ebadat et al.	2013	88	Ebadat et al. (2013)
Varick L. Erickson et al.	2014	80	Erickson et al. (2014)
Luis Miguel et al.	2016	94–98	Candanedo and Feldheim (2016)

indoor temperature from the sensor based data using ELM and OSELM (Kumar et al. 2016). Some of the researchers have presented applications in diverse domains such as intelligent driver drowsiness detection using machine learning approaches on sensor data (Kumar et al. 2018a). Kumar et al. (2018c) proposed a novel method to predict energy load through design and structural attributes and used ELM and OSELM. Researchers also proposed the ensemble model to predict the energy performance of the buildings (Kumar et al. 2018b). Kumar et al. (2019) proposed some models to optimise the number of sensors in the given situation in ambient energy consumption by using metaheuristics approaches. Some researchers developed models from wireless sensor network data and monitored occupancy (Erickson et al. 2014). With the model, the authors reported that it is possible to have annual energy saving up to 42% of energy. A new study Dodier et al. (2006) was proposed to detect the building occupancy using sensor belief networks. In another research work used the Markov based models and reported an accuracy of 80% in the detection of the number of occupants. In such situations, the authors have also proposed models using support vector machine(SVM) and artificial neural networks (ANNs). Zhang et al. (2012) proposed occupancy detection based on the relative information gain (RIG). In this study, the correlation between the number of occupants in the environment was quantified as 77.65% for humidity, 73.42% for acoustics, 67.14% for CO_2 and 37.39% for temperature. Agarwal et al. (2010) proposed pressure sensors based on wireless driven occupancy detection. Tomastik et al. (2010) proposed a real-time prediction model in building occupancy status using video and camera in addition to the passive Infrared sensors.

Some model proposed using a combination of environmental variables such as temperature, humidity, CO_2, light, sound and motion to detect occupancy status and the number of occupants (Yang et al. 2012) and demonstrated accuracy level of 87.62% for self-estimation data and 64.83%. Ekwevugbe et al. (2013) and Beltran et al. (2013) reported the accuracy from 75 to 84.5% with the proposed model. Out of all the work done, some of the researchers stated that CO_2 level, door status and light various contributes significantly to the occupancy status results. Table 1 list some of the recent works proposed models and with accuracy information. After going through the recent work, it is clear that model based on sensors data are cost effective for the occupancy detection in terms of processing of information but getting good accuracy, efficiency, real-time adaptation and generalizations has been a challenge. This paper proposes an ensemble model based on the shallow architecture of ELM and OSELM to make them accurate, efficient and adaptive to real-time.

3 Machine learning approaches

3.1 ELM

Extreme learning machine (ELM) (Huang et al. 2006, 2015) is a neural network-based method which has a single layer of hidden nodes layer connected from the input layer through weights. It is one type of ANNs but with some specific features which make it different from other machine learning approaches based on neural networks. Other approaches suffer from parameter tuning, local minima and generalisation problems while ELM uses different approaches which avoid the parameter tuning. It the ELM weights are randomly assigned and fixed. Hidden nodes and outputs linking through weights is a simple linear learning process. ELM is used by many researchers in several types of applications such as classification, regression, time series forecasting and clustering. ELM

gives very high efficiency in terms of the processing of the information as compared to the other machine learning approaches without compromising on accuracy. In the ELM, determination of the optimal number of hidden layer neuron is time-consuming but can be done with the help of kernal based approach in ELM (Huang 2014). Lets Suppose, There are M number of training samples such as $X = (x_{i1}, x_{i2}, \ldots, x_{iM}) \epsilon \Re^{\mathfrak{M}}$ and $\ldots\ldots N$ and their corresponding output values are represented as vector $Y = (y_{i1}, y_{i2}, \ldots, y_{iN})$. Suppose, k represnet the number of neurons and P hidden neurons. Let G be considered a real value function represented as $G(w_i, b_i, x)$ whihc is the output of the ith hidden neuron with respect to the input vector X and biases $b_i \epsilon \Re$. The Bias values b_i is corresponding to the input vector X and weight vector $w_i = (w_{i1}, w_{i2}, \ldots, w_{iM})$. In this concept, w_{ij} is weight of the links connecting the ith input neuron to the jth hidden neuron. The output value function Y can be given as.

$$Y_P(x_j) = \sum_{i=1}^{P} \eta_i G(w_i, b_i, x_j), \quad j = 1, \ldots, M \tag{1}$$

where η, represent the output weight matrix. Equation (1) can be rewritten as

$$\mathbb{Z}\eta = Y \tag{2}$$

where

$$\mathbb{Z}(\hat{w}, \hat{b}, \hat{x}) = \begin{bmatrix} G(w_1, b_1, x_1) & \ldots & G(w_P, b_P, x_1) \\ \vdots & \ddots & \vdots \\ G(w_1, b_1, x_M) & \ldots & G(w_P, b_P, x_M) \end{bmatrix}_{M \times P} \tag{3}$$

$$\text{where} \quad \begin{matrix} \hat{w} = \{w_1, w_2, \ldots, w_P\} \\ \hat{b} = \{b_1, b_2, \ldots, b_P\} \\ \hat{x} = \{x_1, x_2, \ldots, x_M\} \end{matrix}$$

$$\eta = \begin{bmatrix} \eta_1^T \\ \eta_2^T \\ \vdots \\ \eta_P^T \end{bmatrix}_{P \times N} \quad and \quad Y = \begin{bmatrix} y_1^T \\ y_2^T \\ \vdots \\ y_M^T \end{bmatrix}_{M \times N} \tag{4}$$

In the above mathematical $\mathbb{Z}$ is hidden layer output matrix of (SLFNs) with ith column of $\mathbb{Z}$ being the ith hidden neuron output with respect to the input $\hat{x}$. After the assignment of the random weight and fixing them, Eq. 2 transform itself into linear equation. Weights of the ouweigh matrix can be determined as follows.

$$\eta = \mathbb{Z}^{\dagger} Y. \tag{5}$$

The inverse of the matrix is different here which is called $\mathbb{Z}^{\dagger}$, Moore-Penrose generalised inverse (Bishop 2007) and can calculated if the $\mathbb{Z}^T\mathbb{Z}$ is non singular as follows $\mathbb{Z}^{\dagger} = (\mathbb{Z}^T\mathbb{Z})^{-1}\mathbb{Z}^T$. The activation function in additive nodes, $g(x) : \Re \rightarrow \Re$ can be any activation function. Lets us suppose it is radial bias function (rbf), sigmoid function (sig), sinusoid function (sin), then $G(w_i, b_i, x)$ is given by $G(w_i, b_i, x) = g(w_i \cdot x + b_i), b_i \epsilon \Re$ Here, $w_i \cdot x$ denotes the inner product of the vector w_i and x in $\Re^n$.

3.2 Online sequential ELM (OSELM)

In the real world circumstances, sometimes complete training data is not available. Data comes in streams and machine has to learn from the recent data only to do the classification, regression and clustering. ELM is batch-oriented machine learning approach which works on static data. Liang et al. (2006), Lan et al. (2009) introduced online sequential ELM(OSELM) which is the improvement over ELM for real-time data adaptability. This algorithm accepts data in the chunks of the equal or varying sizes and learning from the very recent data fetched (Liang et al. 2006; Lan et al. 2009).

More specifically, OSELM can learn from the very less and recently fetched training data without looking into historic training data. Additionally, OSELM does not take into account the samples which have been used in the training process already. Similarly, It also takes inout wrghts and biases randomly and fixed them to get the output matrix. But here the samples are in the form of chunks then next chunk like this. The output weight matrix η is caclulated by using recursive least squares algorithm (Huang 2014) and given as

$$\tilde{\eta} = \mathbb{Z}^{\dagger} Y \tag{6}$$

where $rank(\mathbb{Z}) = P$, P is the number of hidden neurons. Here OSELM has two phases. One of initialization phase and other is a sequential phase. The first chunk helps in the initialization of the parameters and weights and then the sequential solution of Eq. (6) results in implementing OSELM. The learning process is as follows. Let us suppose, given training sample $\mathbb{N} = \{(x_i, y_i) \| x_i \epsilon \Re^{\mathfrak{M}}, y_i \epsilon \Re^{\mathfrak{N}}, \mathfrak{i} = , \ldots, \mathfrak{N}\}$ where N training data samples are available initially (Liang et al. 2006). In the initailization phase, OSELM take a chunk of training samples $\mathbb{N}_0 = \{(x_i, y_i) \| i = 1, \ldots, N_0\}$ from the given training set $\mathbb{N}$ where $N_0 \geq P$. After this, randomly generate the hidden neuron parameters $(w_i, b_i) i = 1, \ldots, P$ and fix them to calculate the hidden layer output matrix $\mathbb{Z}_0$ as given below.

$$\mathbb{Z}_0 = \begin{bmatrix} G(w_1 b_1 x_1) & \cdots & G(w_P b_P x_1) \\ \vdots & \ddots & \vdots \\ G(w_1 b_1 x_{N_0}) & \cdots & G(w_P b_P x_{N_0}) \end{bmatrix}_{N_0 \times P} \tag{7}$$

After this, initial output weight matrix $\eta^{(0)}$ values are estimated and given as

$$\eta^{(0)} = \mathbb{Z}_0^{\dagger} Y_0 \tag{8}$$

where $\mathbb{Z}_0$ and Y_0 is defined as $\mathbb{Z}_0^{\dagger} = (\mathbb{Z}_0^T \mathbb{Z}_0)^{-1} \mathbb{Z}_0^T$, $\mathbb{G}_0 = \mathbb{Z}_0^T \mathbb{Z}_0$ and $Y_0 = \{y_1, \ldots, y_{N_0}\}^T$.

Value of k is set to zero. This process initialize the parameters after fist chunk of data. In the sequential learning phase, next chunk of training sample is taken, let us suppose it is $(k+1)$th of the training data. Then next chunk samples are represented as

$$\mathbb{N}_{k+1} = \{(x_i, y_i) | i = \left(\sum_{j=0}^{k} N_j\right) + 1, \ldots, \left(\sum_{j=0}^{k+1} N_j\right) \tag{9}$$

where each chunk might be of varying sizes $N_0, N_1, \ldots, N_{k+1}$. Value in the vector $Y_{(k+1)}$ are given as

$$Y_{(k+1)} = \begin{bmatrix} y^T_{(\sum_{j=0}^{k} N_j + 1)} \\ \vdots \\ y^T_{(\sum_{j=0}^{(k+1)} N_j)} \end{bmatrix}_{N_{(k+1)} \times N} \tag{10}$$

. Next step is the computation of the value of the $\mathbb{Z}_{(k+1)}$ as given

$$\begin{bmatrix} G(w_1, b_1, x_{(\sum_{j=0}^{k} N_j + 1)}) & \cdots & G(w_P, b_P, x_{(\sum_{j=0}^{k} N_j + 1)}) \\ \vdots & \ddots & \vdots \\ G(w_1, b_1, x_{(\sum_{j=0}^{k} N_j + 1)}) & \cdots & G(w_P, b_P, x_{(\sum_{j=0}^{k} N_j + 1)}) \end{bmatrix}_{N_0 \times P} \tag{11}$$

After this, the $(k+1)$th output weight matrix $\eta^{(k+1)}$ is determined as

$$\eta^{(k+1)} = \eta^{(k)} + \mathbb{G}_{(k+1)}^{-1} \mathbb{Z}_{(k+1)}^T (Y_{(k+1)} - \mathbb{Z}_{(k+1)} \eta^{(k)}) \tag{12}$$

where $\mathbb{G}_{k+1} = \mathbb{G}_k + \mathbb{Z}_{k+1}^T \mathbb{Z}_{k+1}$. Now the value of k is increased by one and control goes back to start of sequential phase.

4 Methodology

This section deals with the process of the experiments conducted according to the aim of the paper.

4.1 Data description

Basically, in this research experiments, data consist of five attributes which are used to classify the status of occupancy in the room. Main attributes are temperature, humidity, light, humidity ratio and level of CO_2. All these parameters together obtained with the help of sensors and occupancy of people of the room also recorded and placed on the UCI machine learning repository (Candanedo 2016). Insights about data attribute pattern can be obtained within details with Figs. 1 and 2 which displays the plot with data and time. Training and testing division of the data is 75% and 25%.

4.2 Data processing

After obtaining data from the UCI machine learning repository (Athanasios and Angeliki 2012), It is cleaned and prepossessed. Data Cleaning deals with data scrubbing, detection of error and their removal. This improves the data quality. Absent values are placed with the averages of the values in our experiment. It has been proved that machine learning algorithms work better with normalized data set which ensure the standard features values, removing the possibility of biases and lopsided learning and removing possibility of unfairer representation. Normalization has been performed using following formula.

4.3 Evaluation determinants

Evaluation and its comparison with other models of needs one determinants of accuracy. In the classification one such measure is accuracy defined as follows.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{13}$$

where TP represents true positive, TN represents true negative, FP represents false positive and FN represent false negative.

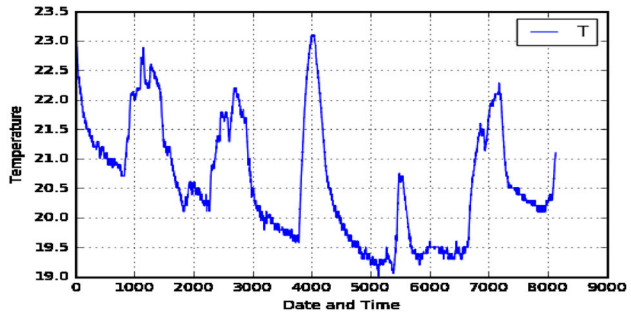
5 Proposed model description

5.1 Statistical Analysis

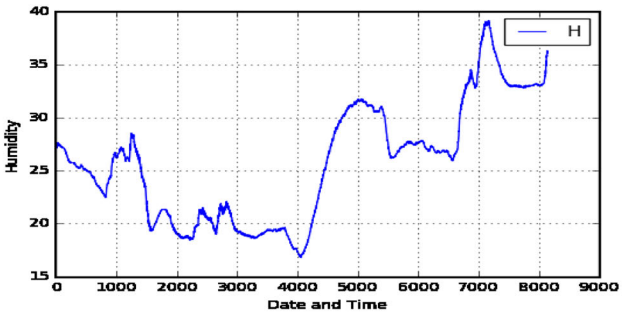
In our experiment, data is statistically analyzed in this section. Data has been plotted across the date and time feature. Histogram and scatter plot is depicted in Figs. 1 and 2. Observing histogram gives the insight about the features of the dataset that some of the features show Gaussian or normal distribution probability densities. The Gaussian nature helps in placing the model and tuning the kernel functions for neural network-based machine learning model. Tables 2 and 3 show co variance and p values for the features. Table 2 displays the relationship of features with target parameter and among themselves. Temperature and light variable are closely related while humidity and temperature are inversely correlated (Fig. 3).

5.2 Iterative ensemble model

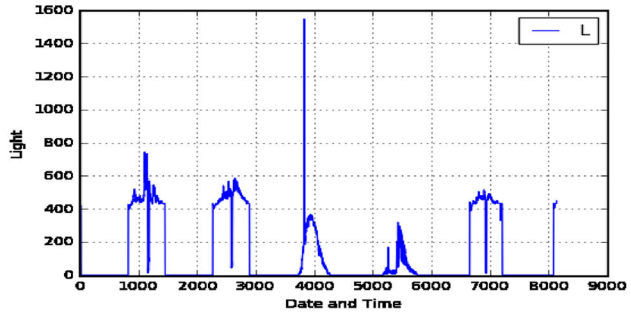
Classification process is the statistical technique which exploits the relationship of some independent attributes with the dependent attribute called class or category and



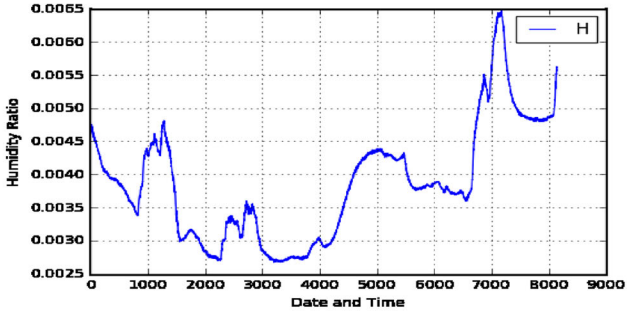
(a) Temperature pictorial representation



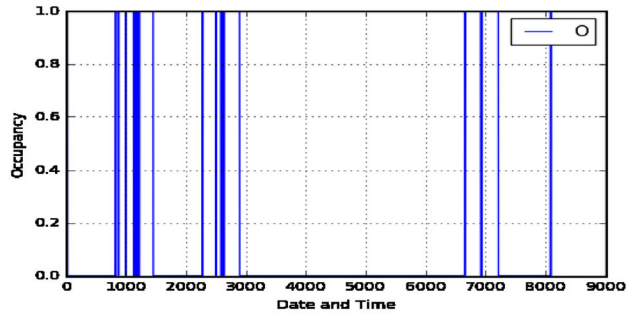
(b) Humanity pictorial representation



(c) Light pictorial representation



(d) Humidity ratio pictorial representation



(e) Occupancy pictorial representation

Fig. 1 Plot of all the features used to predict occupancy

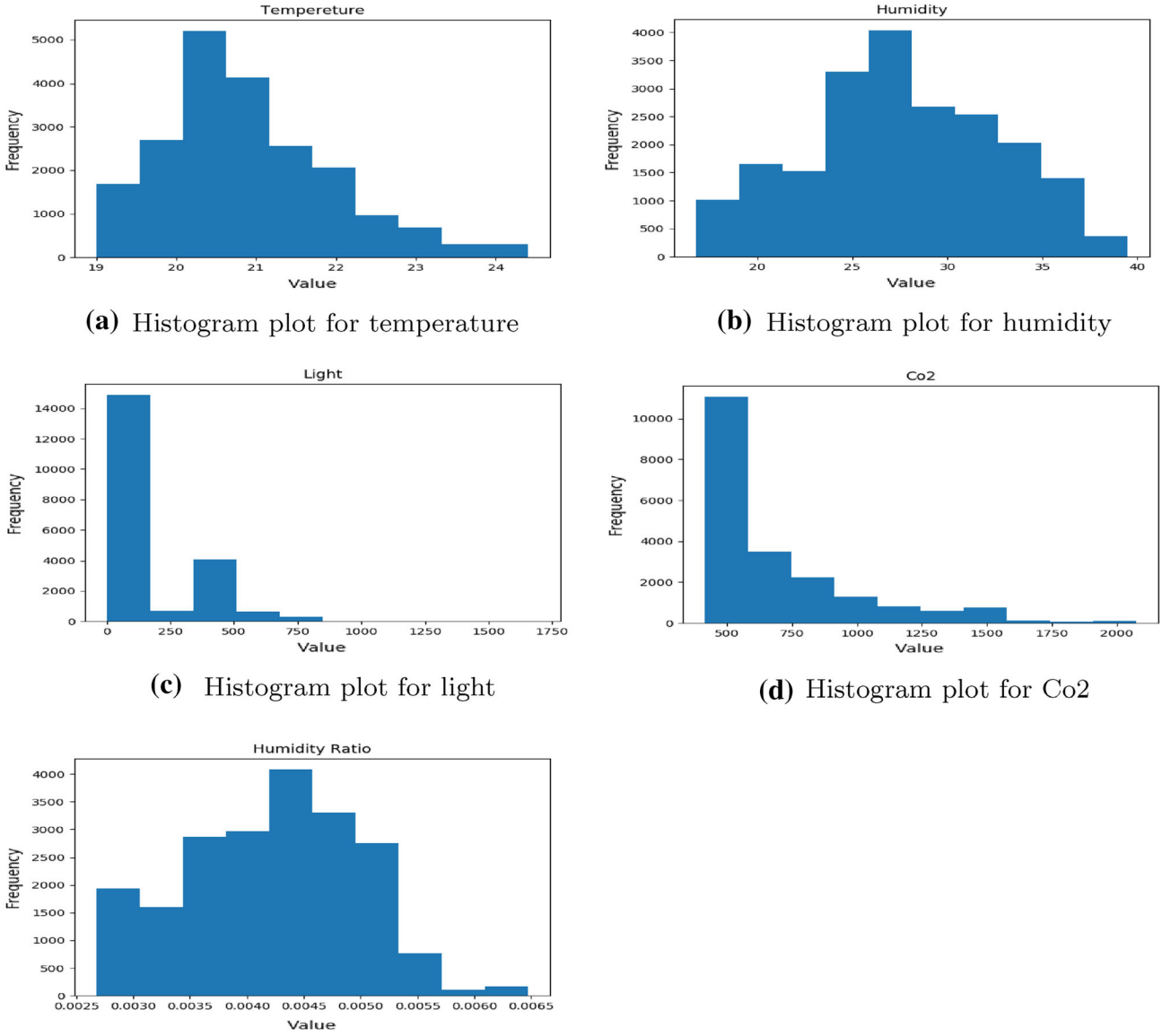
tries to predict the class with the relationship of independent variables. For instance, in this experiment, we have total 7 variables in the data set. Out of them six variables $X_1, X_2, \ldots X_6$ contribute in some way to variable Y which is class that represent whether room is occupied or not in binary form. From the experiments, Each variables from X_1 to X_6 may have their own correlation value with class item occupied or not that can be estimated with various classifiers algorithms (Tables 4, 5). So mathematically classification can be defined as

$$O/E = F(X_1 \cdots X_6) \tag{14}$$

where O and E represent room occupied or empty respectively

$$X_n = (x - x_{max})/(x_{max} - x_{min}) \tag{15}$$

and X_1 to X_7 are the attributes or features such as date and time, temperature, humidity, light, CO_2 and humidity ratio. A better correlation function helps in a good level of accuracy. In our experiment, we are using all the features to predict the room occupancy. Proposed model phases are shown in Fig. 4. This model accepts the data and performed cleaning operation by removing the wrong data and place average value at the missing data points. After cleaning, data normalisation is performed. This experiment is conducted with two type of data with normalisation and without normalisation. After data normalisation, data is divided into two segments for training and testing purpose



(a) Histogram plot for temperature

(b) Histogram plot for humidity

(c) Histogram plot for light

(d) Histogram plot for Co2

(e) Histogram plot for humidity ratio

Fig. 2 Histogram for all the features used to predict occupancy

Table 2 Co-variance with features

Features	Temperature	Humidity	Light	CO_2	Humidity ratio	Occupancy
Temperature	1.0000	− 0.4729	0.7055	0.1996	− 0.0104	0.5218
Humidity	− 0.4729	1.0000	− 0.1875	−0.0792	0.8841	− 0.0878
Light	0.7055	− 0.1875	1.0000	0.1902	0.1627	0.9145
CO_2	0.1996	− 0.0792	0.1902	1.0000	0.0432	0.2619
Humidity ratio	− 0.0104	0.8841	0.1627	0.0432	1.0000	0.1902
Occupancy	0.5218	− 0.0878	0.9145	0.2619	0.1902	1.0000

with 75% and 25% share. Then it is passed to the machine learning model based on ensemble ELM and OSELM depending on the requirement of the algorithm. Now we get results and note down the maximum accuracy for the experiments and compare the value with literature survey with the same set of parameters that is shown in Tables 1 and 2. In terms of processing, Data is fetched from sources and it is sent to data processing module as shown in Fig. 4.

Table 3 *P* values for features

Features	Temperature	Humidity	Light	CO_2	Humidity ratio	Occupancy
Temperature	1.0000	0.0000	0.0000	0.0000	0.3041	0.0000
Humidity	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
Light	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
CO_2	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
Humidity ratio	0.3041	0.0000	0.0000	0.0000	1.0000	0.0000
Occupancy	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

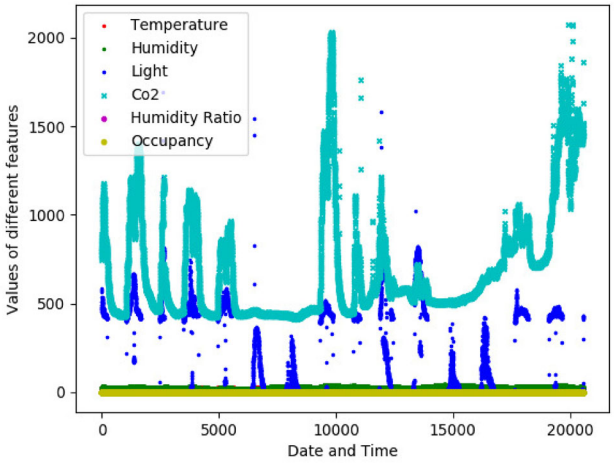


Fig. 3 Scatter plot for all features

There are three steps in this data preprocessing module;data cleaning, data normalisation and statistical analysis. After data cleaning, data is statistically analysed through plotting data features, histograms, covariance, frequency, standard deviation to get insight about the data. All of them are given in Fig. 3, Tables 2, 3 and 4. After this data analysis, Data is sent to proposed ensemble model based on ELM and OSELM with learning mode, activation functions and no of hidden neurons. These activation functions are sigmoid(Sig), radial basis function(RBF) and hard limit(Hardlim). ELM based proposed model works in batch mode and static data while OSELM based model works on dynamic data, real-time data of occupancy detection learning in chunk by chunk of size 10 by 10, 20 by 20 and 30 by 30. After running the model input parameters are selected for optimal results. These parameters are activation function, algorithms, learning mode and no of hidden neurons. These parameters are used to generate the several ensemble outputs through iterative ensemble operation (IEO) in the proposed model. The output obtained with weighted ensemble output predict the occupancy detection with better accuracy.

Table 4 Spearson's rank coeffcient

	Temperature	Humidity	Light	CO_2	Humidity ratio
Occupancy	0.5640	− 0.0828	0.7995	0.3638	0.1648

Table 5 Occupancy detection without normalization

Act function	Algorithm	LM	Average time	Accuracy		Nodes	Candanedo and Feldheim (2016)	
				Training	Testing		CART	LDA
Sigmoid	ELM	Batch	0.1094	98.2369	98.2134	26	94.52	97.9
	OSELM	10 By 10	0.8594	98.6741	98.9284	51	94.52	97.9
		20 By 20	0.5313	98.8831	98.9276	58	94.52	97.9
		[10, 30]	0.1250	98.8374	98.9226	21	94.52	97.9
RBF	ELM	Batch	0.0625	98.2369	98.2134	21	94.52	97.9
	OSELM	10 By 10	2.7500	98.9298	98.5584	127	94.52	97.9
		20 By 20	0.0469	98.2467	98.5034	28	94.52	97.9
		[10, 30]	0.0625	98.2369	98.2134	21	94.52	97.9
Hardlim	ELM	Batch	0.0625	98.2369	98.2134	26	94.52	97.9
	OSELM	10 By 10	0.1563	98.1652	98.2173	7	94.52	97.9
		20 By 20	0.1250	98.1685	98.2148	14	94.52	97.9
		[10, 30]	0.1094	98.1599	98.2167	14	94.52	97.9

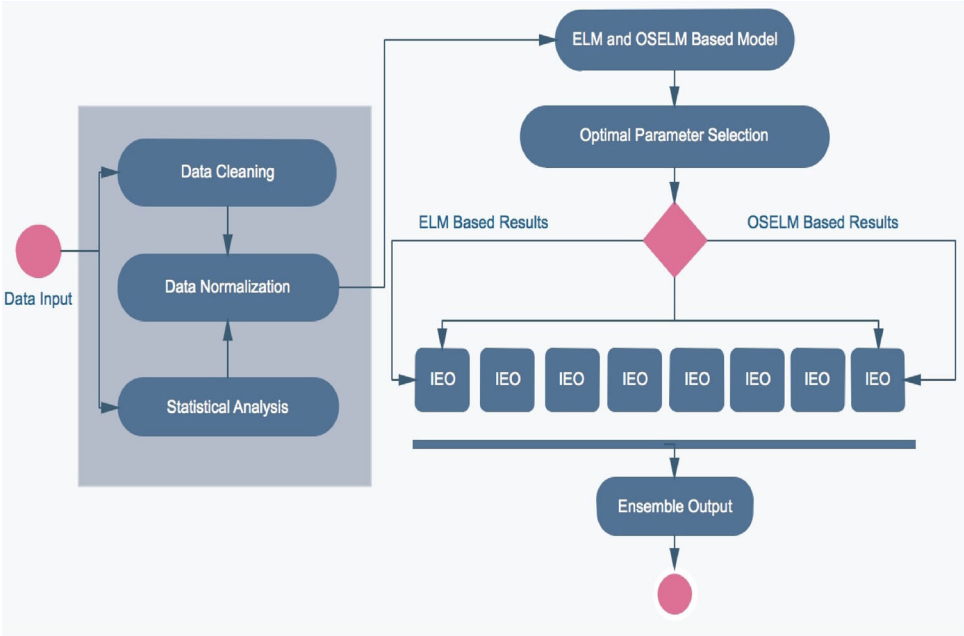


Fig. 4 Iterative ensemble model based on ELM and OSELM

6 Experimental results

Proposed models Iterative Ensemble based on ELM and OSELM are trained and then tested on the data set. The experiment has been conducted with the attributes temperature, humidity, light, the level of CO_2 and humidity ratio. Obtained results were organized in tabular form with activation function, used algorithms, processing time combined training time and testing time and the number of hidden nodes used in the experiments. In table also results of the Iterative Ensemble Based ELM and OSELM model are compared with Candanedo and Feldheim (2016) with the same set of parameter. We discuss the results case by case. The Case 1 compares and discusses the results of Iterative Ensemble model based ELM with literature work discussed in Table 1. The Case 2 discusses the Iterative Ensemble model Based on OSELM on accuracy and time parameters. The Case 3 provides Intra comparison between Iterative Ensemble model based on ELM and OSELM.

Case 1 Proposed model when applied with three activation function and learning mode and hidden neutrons gives good result and improvement in accuracy from 94.52 to 98.21% with normalization of the data. If normalized data is passed, it produces improvement in the results from 98.23 to 98.98%. From the experiment, it is clear that all the activation functions sigmoid, RBF and Hardlim gives the same type of level of results although Hardlim and RBF are slightly better. When processing time is compared it is observed that processing time is fast on non-normalized data while normalized data set takes 2 to 3 times more time in processing. When data is not normalized we need more hidden neutrons to get the optimum level of accuracy whole in case of normalized data set no of hidden neutrons are less than double digit.

Case 2 Proposed model when applied with OSELM and three activation function and learning in batches on real time data coming, with normalization and without normalization of data set provide the same level of accuracy results and improvement of 2–5% error reduction from literature work mentioned in Tables 1 and 6. Real time data is processed and trained in the chunk of 10 by 10, 20 by 20 and any value between 20 and 30. In this situation, in all the learning mode, accuracy is of order 99 percentage but there is a slight improvement when chunk size is 20 by 20 or between 20 or 30. Proposed model based on OSELM performed better in the case of normalized data. Sigmoid activation function based results are better than the results of RBF and hardlim. Processing time is less in learning rate of chunk size 20 by 20 and big chunk she as compared to 10 by 10 which is self-explanatory.

Case 3 In this comparison, proposed model Iterative Ensemble based on ELM provides slightly better accuracy results as compared to proposed Iterative Ensemble model based on OSELM. This may because of the large data available for learning in ELM case although accuracy difference between two models is not significantly big. In normalized data case model based on ELM is better than the model based on OSELM subject to the constrained of the batch. As we increase the batch size the accuracy level of OSELM based model reaches to proposed model based on ELM. Without normalization, iterative ensemble model based on OSELM takes more time as compared to the processing of normalized data. In the case of Iterative Ensemble,

Table 6 Occupancy detection with normalization

Act function	Algorithms	LM	Average time	Accuracy		Nodes	Candanedo and Feldheim (2016)	
				Training	Testing		CART	LDA
Sigmoid	ELM	Batch	3.7860	98.8644	98.9977	7	94.52	97.9
	OSELM	10 By 10	0.5938	98.4457	98.9167	64	94.52	97.9
		20 By 20	0.5625	98.4493	98.8693	81	94.52	97.9
		[10, 30]	0.4063	98.5508	98.8245	82	94.52	97.9
RBF	ELM	Batch	3.0890	98.9968	98.9929	4	94.52	97.9
	OSELM	10 By 10	1.5625	98.4638	98.9407	66	94.52	97.9
		20 By 20	0.9844	98.4404	98.8979	73	94.52	97.9
		[10, 30]	0.6563	98.5067	98.8482	64	94.52	97.9
Hardlim	ELM	Batch	3.8594	98.9993	98.9928	5	94.52	97.9
	OSELM	10 By 10	0.2656	98.0359	98.2737	10	94.52	97.9
		20 By 20	0.1250	98.1808	98.7156	10	94.52	97.9
		[10, 30]	0.1719	98.1179	98.5126	5	94.52	97.9

ELM model takes more time in normalized data set than the time taken without normalized data set. In this situation, the time taken in the normalized situation is many times.

After observing the results obtained, it is clear that proposed model based on ELM and OSELM is better in terms of accuracy than Candanedo and Feldheim (2016) which is based on decision tree (CART) and linear discriminate analysis(LDA). If all the parameters are taken and CART model is applied, It provides accuracy of level 94.52% and LDA provided 97.9% which is close to 98%. But proposed model based on ensemble ELM and OSELM provide two improvements over the CART and LDA. The one is that proposed model allows real time learning and real-time occupancy detection with better accuracy of 98.21 over LDA and CART.

7 Conclusion

In this experiment, we implemented the task of occupancy detection based on environmental attributes obtained from experiment. These attributes included date and time, temperature, humidity, light, the level of CO_2 and humidity ratio. Experiments developed the Iterative Ensemble based model with emerging machine learning techniques called ELM and its variants for the real time learning OSELM. Experiments resulted in the better performance and improved accuracy as compared to the previous methods discussed in literature survey section and Table 1. After the results and their discussion, It is clear that model Iterative ensemble model based on ELM and OSELM performs better with the significance level of generalized accuracy up to 99.2 from 97.3 to 94.23% obtained from CART and LDA. Experiment shows that normalized data produces better accuracy and performance in model. The proposed model is better candidate for implementing in real life scenarios as processing is very fast because of inherent show architecture of ELM and OSELM. For future work, other methods such deep learning, feature extraction and generation can also be used further to improve accuracy.

References

Agarwal Y, Balaji B, Gupta R, Lyles J, Wei M, Weng T (2010) Occupancy-driven energy management for smart building automation. In: Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building, BuildSys'10. ACM, New York, pp 1–6

Athanasios T, Angeliki X (2012) Energy efficiency data set. http://archive.ics.uci.edu/ml/datasets/Energy+efficiency?ref=datanews.io

Beltran A, Erickson VL, Cerpa AE (2013) Thermosense: occupancy thermal based sensing for HVAC control. In: Proceedings of the 5th ACM workshop on embedded systems for energy-efficient buildings (BuildSys'13). ACM, New York, pp 11:1–11:8

Bhattacharya M, Reddy Paramati S, Ozturk I, Bhattacharya S (2016) The effect of renewable energy consumption on economic growth: evidence from top 38 countries. Appl Energy 162:733–741

Bishop CM (2006) Pattern recognition and machine learning. Springer, New York

Brooks J, Goyal S, Subramany R, Lin Y, Middelkoop T, Arpan L, Carloni L, Barooah P (2014) An experimental investigation of occupancy-based energy-efficient control of commercial building indoor climate. In: 53rd IEEE conference on decision and control. IEEE, pp 5680–5685

Brooks J, Kumar S, Goyal S, Subramany R, Barooah P (2015) Energy-efficient control of under-actuated HVAC zones in commercial buildings. Energy Build 93:160–168

Candanedo L (2016) Occupancy detection data set. https://archive.ics.uci.edu/ml/datasets/Occupancy+Detection+

Candanedo LM, Feldheim V (2016) Accurate occupancy detection of an office room from light, temperature, humidity and CO_2 measurements using statistical learning models. Energy Build 112:28–39

Ding S, Zhao H, Zhang Y, Xu X, Nie R (2015) Extreme learning machine: algorithm, theory and applications. Artif Intell Rev 44(1):103–115

Dodier RH, Henze GP, Tiller DK, Guo X (2006) Building occupancy detection through sensor belief networks. Energy Build 38(9):1033–1043

Dong B (2009) Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings

Dong B, Andrews B, Lam KP, Höynck M, Zhang R, Chiou Y-S, Benitez D (2010) An information technology enabled sustainability test-bed (itest) for occupancy detection through an environmental sensing network. Energy Build 42(7):1038–1046

Ebadat A, Bottegal G, Varagnolo D, Wahlberg B, Johansson KH (2013) Estimation of building occupancy levels through environmental signals deconvolution. In: Proceedings of the 5th ACM workshop on embedded systems for energy-efficient buildings (BuildSys'13). ACM, New York, pp 8:1–8:8

Ekwevugbe T, Brown N, Pakka V (2013) Real-time building occupancy sensing for supporting demand driven HVAC operations. In: 7th IEEE international conference on IEEE, digital ecosystems and technologies (DEST), pp 114–119

Erickson VL, Lin Y, Kamthe A, Brahme R, Surana A, Cerpa AE, Sohn MD, Narayanan S (2009) Energy efficient building environment control strategies using real-time occupancy measurements. In: Proceedings of the first ACM workshop on embedded sensing systems for energy-efficiency in buildings (BuildSys'09). ACM, New York, pp 19–24

Erickson VL, Carreira-Perpiñán MÁ, Cerpa AE (2011) Observe: occupancy-based system for efficient reduction of HVAC energy. In: 10th Information processing in sensor networks (IPSN). IEEE

Erickson VL, Carreira-Perpiñán MÁ, Cerpa AE (2014) Occupancy modeling and prediction for building energy management. ACM Trans Sens Netw 10(3):42:1–42:28

Huang G-B (2014) An insight into extreme learning machines: random neurons, random features and kernels. Cogn Comput 6(3):376–390

Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1):489–501

Huang G, Huang G-B, Song S, You K (2015) Trends in extreme learning machines: a review. Neural Netw 61:32–48

Kéry M, Gardner B, Monnerat C (2010) Predicting species distributions from checklist data using site-occupancy models. J Biogeogr 37(10):1851–1862

Kleiminger W, Beckel C, Staake T, Santini S (2013) Occupancy detection from electricity consumption data. In: Proceedings of the 5th ACM workshop on embedded systems for energy-efficient buildings (BuildSys'13). ACM, New York, pp 10:1–10:8

Kumar S, Pal SK, Singh R (2016) Intelligent energy conservation: indoor temperature forecasting with extreme learning machine. In: Proceedings of intelligent systems technologies and applications, vol 2. Springer, Switzerland, pp 977–988

Kumar S, Kalia A, Sharma A (2018a) Predictive analysis of alertness related features for driver drowsiness detection. Adv Intell Syst Comput 736:368–377

Kumar S, Pal KS, Singh R (2018b) Intra elm variants ensemble based model to predict energy performance in residential buildings. Sustain Energy Grids Netw 16:177–187

Kumar S, Pal SK, Singh R (2018c) A novel method based on extreme learning machine to predict heating and cooling load through design and structural attributes. Energy Build 176:275–286

Kumar S, Singh R, Pal SK (2018d) A conceptual architectural design for intelligent health information system: Case study on India. In: Quality, vol 1. IT and business operations: Springer proceedings in business and economics. Springer, Singapore, pp 1–15

Kumar S, Pal SK, Singh R (2019) A novel hybrid model based on particle swarm optimisation and extreme learning machine for short-term temperature prediction using ambient sensors. Sustain Cities Soc 49:101601

Labeodan T, Zeiler W, Boxem G, Zhao Y (2015) Occupancy measurement in commercial office buildings for demand-driven control applicationsâa survey and detection system evaluation. Energy Build 93:303–314

Lamnabhi-Lagarrigue F, Annaswamy A, Engell S, Isaksson A, Khargonekar P, Murray RM, Nijmeijer H, Samad T, Tilbury D, Van den HP (2017) Systems & control for the future of humanity, research agenda: current and future roles, impact and grand challenges. Annu Rev Control 43:1–64

Lan Y, Soh YC, Huang G-B (2009) Ensemble of online sequential extreme learning machine. Neurocomputing 72(13):3391–3395

Liang N-Y, Huang G-B, Saratchandran P, Sundararajan N (2006) A fast and accurate online sequential learning algorithm for feedforward networks. IEEE Trans Neural Netw 17(6):1411–1423

Meyn S, Surana A, Lin Y, Oggianu SM, Narayanan S, Frewen TA (2009) A sensor-utility-network method for estimation of occupancy in buildings. In: Proceedings of the 48th IEEE conference on decision and control/28th Chinese control conference(CDC/CCC 2009). IEEE, Shanghai, pp 1494–1500

Nguyen TA, Aiello M (2013) Energy intelligent buildings based on user activity: a survey. Energy Build 56:244–257

Pérez-Lombard L, Ortiz J, Pout C (2008) A review on buildings energy consumption information. Energy Build 40(3):394–398

Richardson I, Thomson M, Infield D (2008) A high-resolution domestic building occupancy model for energy demand simulations. Energy Build 40(8):1560–1566

Tomastik R, Narayanan S, Banaszuk A, Meyn S (2010) Model-based real-time estimation of building occupancy during emergency egress. Springer, Berlin, pp 215–224

Wang J, Tse N, Poon T, Chan J (2018) A practical multi-sensor cooling demand estimation approach based on visual, indoor and outdoor information sensing. Sensors 18(11):3591

Yang Z, Li N, Becerik-Gerber B, Orosz M (2012) A multi-sensor based occupancy estimation model for supporting demand driven HVAC operations. In: Proceedings of the 2012 symposium on simulation for architecture and urban design (SimAUD'12). Society for Computer Simulation International, pp 2:1–2:8

Yang Z, Li N, Becerik-Gerber B, Orosz M (2014) A systematic approach to occupancy modeling in ambient sensor-rich buildings. Simulation 90(8):960–977

Yang J, Santamouris M, Lee SE (2016) Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings. Energy Build 121:344–349

Zhang R, Lam KP, Chiou Y-S, Dong B (2012) Information-theoretic environment features selection for occupancy detection in open office spaces. Build Simul 5(2):179–188

Zheng X, Wei C, Qin P, Guo J, Yihua Y, Song F, Chen Z (2014) Characteristics of residential energy consumption in China: findings from a household survey. Energy Policy 75:126–135

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of ELECTRONIC MATERIALS
https://doi.org/10.1007/s11664-019-07914-z
© 2020 The Minerals, Metals & Materials Society

EPR and Optical Properties of Green Emitting Mn Activated $Sr_2ZnSi_2O_7$ Phosphors Prepared by Sol–Gel Method

VIJAY SINGH,[1,6] NISHA DEOPA,[2] SUMANDEEP KAUR,[3] A.S. RAO,[3] J.L. RAO,[4] and G. LAKSHMINARAYANA[5]

1.—Department of Chemical Engineering, Konkuk University, Seoul 05029, Republic of Korea. 2.—Department of Physics, Chaudhary Ranbir Singh University, Rohtak Bypass Road, Jind, Haryana 126 102, India. 3.—Department of Applied Physics, Delhi Technological University, Bawana Road, New Delhi 110 042, India. 4.—Department of Physics, S.V. University, Tirupati, Andhra Pradesh 517 502, India. 5.—Intelligent Construction Automation Center, Kyungpook National University, 80, Daehak-ro, Buk-gu, Daegu 41566, Republic of Korea. 6.—e-mail: vijayjiin2006@yahoo.com

A series of Mn^{2+} activated $Sr_2ZnSi_2O_7$ phosphors with variable Mn^{2+} ions concentration were prepared via citrate sol–gel method. The systematic analysis of structure and luminescent properties were carried out by recording x-ray diffraction patterns, scanning electron micrographs, excitation and emission spectra along with electron paramagnetic resonance (EPR) spectra. The optimization of Mn^{2+} ion concentration in the synthesized phosphors was done by analyzing emission spectra. The Mn^{2+} activated $Sr_2ZnSi_2O_7$ phosphors exhibit green emission under 263 nm excitation. The EPR spectra exhibit sextet hyperfine lines with resonance at $g \approx 2.0$ indicating tetrahedral symmetry and covalent bonding between Mn^{2+} ions and O^{2-} ligand ions bound to it. All the aforesaid results support the suitability of Mn^{2+} activated $Sr_2ZnSi_2O_7$ phosphors as visible green emitters in optoelectronic display devices.

Key words: Sol–gel, EPR, Mn^{2+} ions, $Sr_2ZnSi_2O_7$, photoluminescence

INTRODUCTION

It is well known that luminescence plays a major technological role for mankind. The phosphor materials doped with rare earth and transition metal ions have been regarded as the important luminescence sources. Currently, new phosphor materials doped with rare earth and transition metal ions have been investigated and developed largely due to their potential applications in many fields, such as field emission displays, light-emitting diodes (LEDs), cathode ray tubes, x-ray imaging scintillators, and plasma display panels.[1–5]

In recent years, alkaline earth silicates doped with rare earth and transition metal ions have attracted considerable research interests in the field of photoluminescence due to their high luminescent performance and chemical stability. $A_2BSi_2O_7$ (A = Ca, Sr, Ba, B = Mg, Zn) has akermanite- type and melilite-type crystal structures and due to that has gained much research attention.[6,7] The most common crystal structure is the tetragonal akermanite-type.

Currently, alkaline earth silicates $A_2BSi_2O_7$ (A = Ca, Sr, Ba, B = Mg, Zn) co-doped with rare earth ions and transition metal ions have attracted much research interests due to their excellent thermal and chemical stability. Yao et al.[8] reported luminescent properties of $Sr_2ZnSi_2O_7$:Eu^{2+} phosphors prepared by the combustion-assisted synthesis method. Electrical and thermal properties of PTFE-$Sr_2ZnSi_2O_7$ composites and dielectric, thermal and mechanical properties of $Sr_2ZnSi_2O_7$ based polymer/ceramic composites were reported by Joseph et al.[9,10] Hao et al.[11] reported synthesis and photoluminescence of new phosphors M_2(Mg,

(Received September 12, 2019; accepted December 17, 2019)

Published online: 06 January 2020

Zn)Si_2O_7:Mn^{2+} (M = Ca, Sr, Ba). Wang et al.[12] reported crystal size dependence of the persistent phosphorescence in $Sr_2ZnSi_2O_7$:Eu^{2+}, Dy^{3+}. Zhang et al.[13] investigated reddish orange long lasting phosphorescence of Sm^{3+} in $Sr_2ZnSi_2O_7$:Sm^{3+} phosphors.

The Mn^{2+} activated inorganic phosphors mostly show broad-band emissions due to the electronic transitions between $^6A_1$ ground state and $^4T_1$ excited state of Mn^{2+} ions. In several inorganic compounds such silicates, aluminates and sulfides the luminescence from Mn^{2+} plays an important role where it emits in the visible region. Mn^{2+} ions show green, orange, and red photoluminescence regions under UV irradiation depending mostly on the host matrix because of the sensitivity of the d–d transition in Mn^{2+} to the crystal field and, therefore, widely used as phosphors for lamps and displays.[14,15] In recent years, several research groups have done considerable efforts towards the synthesis and characterization of various Mn^{2+} activated host materials. Singh et al.[16] reported luminescence and EPR study of Mn^{2+} activated $MgSrAl_{10}O_{17}$ green-emitting phosphor. Tripathi et al.[17] reported the effects of Mn^{2+} in CdS nanocrystals. Li et al.[18] reported Mn^{2+} activated MgAlON a green-emitting transparent ceramic phosphor for high-power white LED. Singh et al.[19] investigated EPR and luminescence properties of combustion synthesized $LiAl_5O_8$:Mn phosphors. More recently, Li et al.[20] reported structural, persistent luminescence properties and trap characteristics of an orthosilicate Mn^{2+} doped $LiGaSiO_4$ phosphor.

The electron paramagnetic resonance (EPR) is a very important technique which provides information on the nature of electron spins in the systems. EPR investigations of Mn^{2+} ions in the host lattices is interesting and have received much attention due to the EPR parameters of Mn^{2+} ion are very sensitive to the local symmetry. The main advantage of using Mn^{2+} ion as the spin probe is that its EPR spectrum can easily be recorded at room temperature and this spectrum provides useful information about the oxidation state and site occupancy of these ions in the host structure. As a part of our interest on EPR spectroscopy, in recent years we have done EPR analysis on several Mn^{2+} activated phosphor materials.[21,22] In continuation of our interest, in the present report, we have done EPR analysis on Mn^{2+} activated $Sr_2ZnSi_2O_7$ phosphor material. The phosphor material was prepared by citric-acid based sol–gel method and characterized by using x-ray diffraction (XRD), scanning electron microscope (SEM), photoluminescence (PL) and EPR spectroscopy.

MATERIALS PREPARATION AND ANALYSIS

A series of $Sr_{2-x}ZnSi_2O_7$:xMn (x = 0.01, 0.03, 0.05, 0.07, 0.09 and 0.11) samples were prepared via the citrate sol–gel method by considering the stoichiometric amounts of high-purity $Sr(NO_3)_2$ (A.R.) $Zn(NO_3)_2{\cdot}6H_2O$ (A.R.), SiO_2, citric acid and $Mn(NO_3)_2{\cdot}4H_2O$. The details of sample composition and starting materials are given in Table I. Initially the stoichiometric amounts of Strontium nitrate, Zinc nitrate, Fumed silica, and Manganese nitrate were dissolved in de-ionized water and ethanol solution in the volumetric ratio of 1:4 under stirring. After 10 min, 10 mol of citric acid (citric acid/metal ion = 2:1, molar ratio) was added to the above solution and stirring was continued for 1 h to get complete homogeneous solution. Here the citric acid serves as a complexing/chelating agent while stirring for about 1 h to make the complete homogeneous solution of all reactants. The obtained transparent aqueous solution was kept in an oven at 110°C for 12 h to get the dried gel. Then the obtained dried-gel was heated in a furnace at 400°C for 2 h. Finally, the obtained black-colored flakes were ground using mortar and pestle and annealed at 1000°C for 6 h to obtain final samples.

The PXRD (Powder x-ray Diffraction) patterns of the samples were recorded using a RIGAKUM Miniflex-II diffractometer. Cu-Kα radiation (λ = 1.5406Å) has been used as the x-ray source. The XRD patterns were taken with scan rate of 5°/min in the 2θ range of 10°–80°. Morphology details were obtained using a SEM (S-3400, Hitachi, Japan). Photoluminescence measurements were carried out at room temperature on a Shimadzu RF-5301PC, spectrofluorophotometer equipped with Xenon flash lamp. The EPR spectra of the sample were recorded on a JEOL JES-FA200 ESR Spectrometer, operating in the X-band frequencies, with a field modulation of 100 kHz.

RESULTS AND DISCUSSION

Crystal Structure and Particle Size Analysis

In Ref. 23 it is reported that melilite-related compounds have the general formula $X_2T^1T^2A_7$ where X = Na, Ca, Sr, or Ba, T^1 = Be, Mg, Al, Co, or Zn and T^2 = Al, Si or Ge and A is an anion, frequently O, but also F or S. The melilite-type $Sr_2ZnSi_2O_7$ forms tetragonal structure with $P\bar{4}2_1m$ space group. Ardit et al.[23] investigated the $Sr_2ZnSi_2O_7$ with x-ray powder diffraction. The crystal structure of $Sr_2ZnSi_2O_7$ having $P\bar{4}2_1m$ space group and a = 8.0007(1) Å, c = 5.1722(1) Å lattice parameters was refined by Rietveld method. The structure can be explained as layers of linked tetrahedra with Si_2O_7 dimers and four-coordinated Zn cations connected together by Sr cations that are accommodated in a distorted eight coordinated antiprismatic site.[23] The x-ray diffraction patterns of the $Sr_{2-x}ZnSi_2O_7$:xMn (x = 0.01 mol, 0.03 mol, 0.05 mol, 0.07 mol, 0.09 mol and 0.11 mol) synthesized by the sol–gel method at 1000°C temperatures are shown in Fig. 1a. Results show that the samples possess a good crystallinity nature. All the

Table I. Detailed information of sample composition and starting materials

Samples composition	Starting materials				
$Sr_{1.99}ZnSi_2O_7$:0.01Mn	Sr = 0.8421 g	Zn = 0.5948 g	Si = 0.2400 g	C.A = 3.8420 g	Mn = 0.0050 g
$Sr_{1.97}ZnSi_2O_7$:0.03Mn	Sr = 0.8337 g	Zn = 0.5948 g	Si = 0.2400 g	C.A = 3.8420 g	Mn = 0.0150 g
$Sr_{1.95}ZnSi_2O_7$:0.05Mn	Sr = 0.8252 g	Zn = 0.5948 g	Si = 0.2400 g	C.A = 3.8420 g	Mn = 0.0251 g
$Sr_{1.93}ZnSi_2O_7$:0.07Mn	Sr = 0.8167 g	Zn = 0.5948 g	Si = 0.2400 g	C.A = 3.8420 g	Mn = 0.0351 g
$Sr_{1.91}ZnSi_2O_7$:0.09Mn	Sr = 0.8083 g	Zn = 0.5948 g	Si = 0.2400 g	C.A = 3.8420 g	Mn = 0.0451 g
$Sr_{1.89}ZnSi_2O_7$:0.11Mn	Sr = 0.7978 g	Zn = 0.5948 g	Si = 0.2400 g	C.A = 3.8420 g	Mn = 0.0552 g

Sr = $Sr(NO_3)_2$, Zn = $Zn(NO_3)_2{\cdot}6H_2O$, Si = SiO_2, C.A = Citric acid, Mn = $Mn(NO_3)_2{\cdot}4H_2O$.

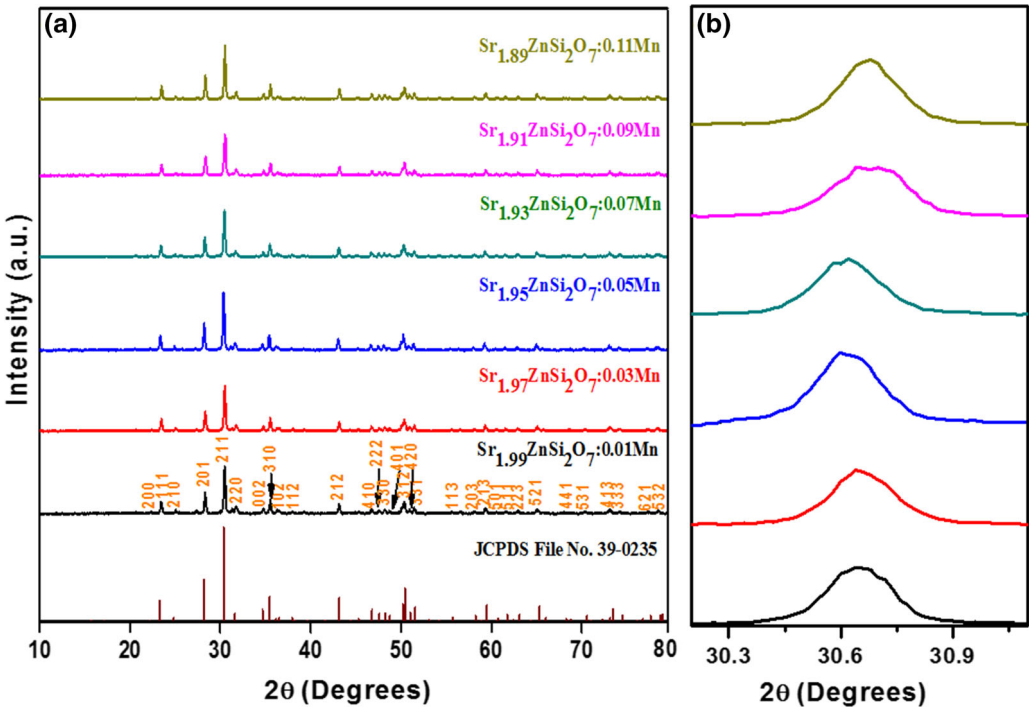


Fig. 1. (a) Powder XRD patterns of $Sr_{2-x}ZnSi_2O_7$:xMn phosphor; (b) magnified images of intense diffraction peak corresponding to 2 1 1 plane for all the samples

synthesized samples agree well with the JCPDS card of $Sr_2ZnSi_2O_7$ (No. 39-0235) confirming the tetragonal structure. The doping of Mn^{2+} ions with a present amount does not destroy the structure of the samples. In Refs. 9–11 and 24 it is shown that there are two different types of tetrahedra in $Sr_2ZnSi_2O_7$ structure. The Zn atoms are in tetrahedra of oxygen, all four of which are shared by adjacent [SiO_4] tetrahedra, which themselves are linked in pairs to form [Si_2O_7] groups. The interconnected tetrahedral units are held together by interlayer Sr^{2+} having eight-fold oxygen coordination. Layers of eight-coordinated Sr run parallel to the $Zn(Si_2O_7)$ sheets perpendicular to the unit cell b axis. In $Sr_2ZnSi_2O_7$ structure, Mn^{2+} or Tb^{3+} can substitute the site of Zn^{2+} or Sr^{2+} and give green emission.[24] It can be expected that doping of another atom into any structure leads to peak shift because atomic radius substitution into an atom vacancy is different. This can also be interpreted from the shift in the main peak of XRD as shown in Fig. 1b. Sr^{2+} ion has an ionic radius of 1.26 Å in a eight-fold coordination. In the same coordination, Mn^{2+} has an ionic radius of 0.96 Å. Zn^{2+} ion has an ionic radius of 0.60 Å in a four-fold coordination. In the same coordination, Mn^{2+} has an ionic radius of 0.66 Å. Thus, doping ions of Mn^{2+} ions are likely to substitute the zinc sites. The crystallite size was calculated from the main diffraction XRD peak by using the well-known Scherer's equation

$$D = 0.9\,\lambda/\beta\cos\theta, \qquad (1)$$

where λ is the wavelength of incident x-ray, θ is the corresponding Bragg's diffraction angle, and β is the

FWHM of the (211) peak. The crystallite size calculated from this method is found to be in the range of 37–45 nm. The calculated crystallite size and FWHM for all samples are given in Table II.

SEM Analysis

The morphology of powders was investigated by SEM, from which the grain size and typical morphologies were obtained. Scanning electron micrographs of the $Sr_{1.91}ZnSi_2O_7$:0.09Mn sample annealed at 1000°C are shown in Fig. 2. From Fig. 2a and b, it is clearly observed that the individual particles are seeming to be sub-micro sized plate-like crystals, and they partially fused to form hard agglomerates. Furthermore, these agglomerates particles are covered with differently shaped smaller particles. We can also observe some pores which are probably formed due to the release of some gaseous byproducts during the process of sintering. From Fig. 2c and d, it is clear that there are many aggregates of small particles present which are in nanosized with irregular shape which is good agreement with calculated crystallite size from XRD data.

Photoluminescence Analysis

The photoluminescence (PL) spectra of Mn^{2+} ions activated $Sr_2ZnSi_2O_7$ phosphors are shown in Fig. 3. From the Fig. 3a it is conspicuous that, the as prepared phosphors are exhibiting one broad excitation band centered at 263 nm under 528 nm emission wavelength. The intense excitation band observed at 263 nm is taken as an excitation wavelength (λ_{exc} = 263 nm) to record the emission spectra of Mn^{2+} ions activated $Sr_2ZnSi_2O_7$ phosphors as shown in Fig. 3b. The emission band at 528 nm in Fig. 3b is attributed to the transition from $^4T_1$ (^4G) excited state to the $^6A_1$ (^6S) ground state. The Mn^{2+} ions in $Sr_2ZnSi_2O_7$ phosphor substitute Zn^{2+} ions, since the ionic radii of Mn^{2+} ions and Zn^{2+} ions are in close proximity. Each Mn^{2+} ion is having four bonds because of the tetrahedral coordination and the tetrahedrally coordinated Mn^{2+} ions are perturbed due to a weak crystal field. This in turn results in green emission from Mn^{2+} ions activated $Sr_2ZnSi_2O_7$ phosphors. The green emission of Mn^{2+} ions activated phosphors may be represented as

$$E(h\upsilon) + Mn^{2+} = Mn^{3+} + e^{-}. \quad (2)$$

The energy E is responsible for photo-excitation of Mn^{2+} ions in $Sr_2ZnSi_2O_7$ phosphor. This in turn excites ground state free electrons to higher states within the forbidden band. Non-radiative decays start from these higher states to $^4T_1$ excited state and from $^4T_1$ state radiative transition takes place towards ground state $^6A_1$ exhibiting green emission.[16] Further, from Fig. 3 it can be seen that the photoluminescence intensity of Mn^{2+} ions activated $Sr_2ZnSi_2O_7$ phosphors increases with increase in Mn^{2+} ion concentration up to 0.09 mol ($Sr_{1.91}ZnSi_2O_7$:0.09Mn) and beyond that it decreases. This may be due to energy transfer processes through cross-relaxation channels and subsequent concentration quenching is observed. The variation in emission intensity with Mn^{2+} ions concentration can be seen from Fig. 4. From Fig. 4, it is observed that, under 263 nm excitation wavelength, the PL intensity is maximum for $Sr_{1.91}ZnSi_2O_7$:0.09Mn phosphor. Therefore, among all the as prepared phosphors, $Sr_{1.91}ZnSi_2O_7$:0.09Mn phosphor is aptly suitable for photonic device application.

The CIE color coordinates exhibited by a material can be evaluated by PL spectra by using the Commission International de I'Eclairage France CIE system.[25] The emission spectra of Mn^{2+} ions activated $Sr_2ZnSi_2O_7$ phosphors spread in the green region of the visible spectrum. Figure 5 shows the CIE color coordinates of Mn^{2+} ions activated $Sr_2ZnSi_2O_7$ phosphors under 263 nm excitation wavelength. The CIE color coordinates of the prepared phosphors are found in the green region of the CIE chromaticity diagram (Table III). It can be seen from Fig. 5, that among all the phosphors, the color co-ordinates evaluated for $Sr_{1.91}ZnSi_2O_7$:0.09Mn phosphor exhibit intense green emission. Therefore, $Sr_{1.91}ZnSi_2O_7$:0.09Mn phosphor is optimum for green color generation. Figure 6 shows the digital photographs of $Sr_{1.91}ZnSi_2O_7$:0.09Mn sample under 254 nm excitation. The luminescence is strong enough to be visible to the eye in dark–light. These results are in consistent with the results possessed by PL spectra of Mn^{2+} ions activated $Sr_2ZnSi_2O_7$ phosphors in Fig. 3.

Table II. FWHM and crystallite size of $Sr_{2-x}ZnSi_2O_7$:xMn phosphor

Sample composition	FWHM (°)	Crystalline size (nm)
$Sr_{1.99}ZnSi_2O_7$:0.01Mn	0.1999	43.06
$Sr_{1.97}ZnSi_2O_7$:0.03Mn	0.2019	42.63
$Sr_{1.95}ZnSi_2O_7$:0.05Mn	0.2018	42.64
$Sr_{1.93}ZnSi_2O_7$:0.07Mn	0.2066	41.66
$Sr_{1.91}ZnSi_2O_7$:0.09Mn	0.2271	37.90
$Sr_{1.89}ZnSi_2O_7$:0.11Mn	0.192	44.83

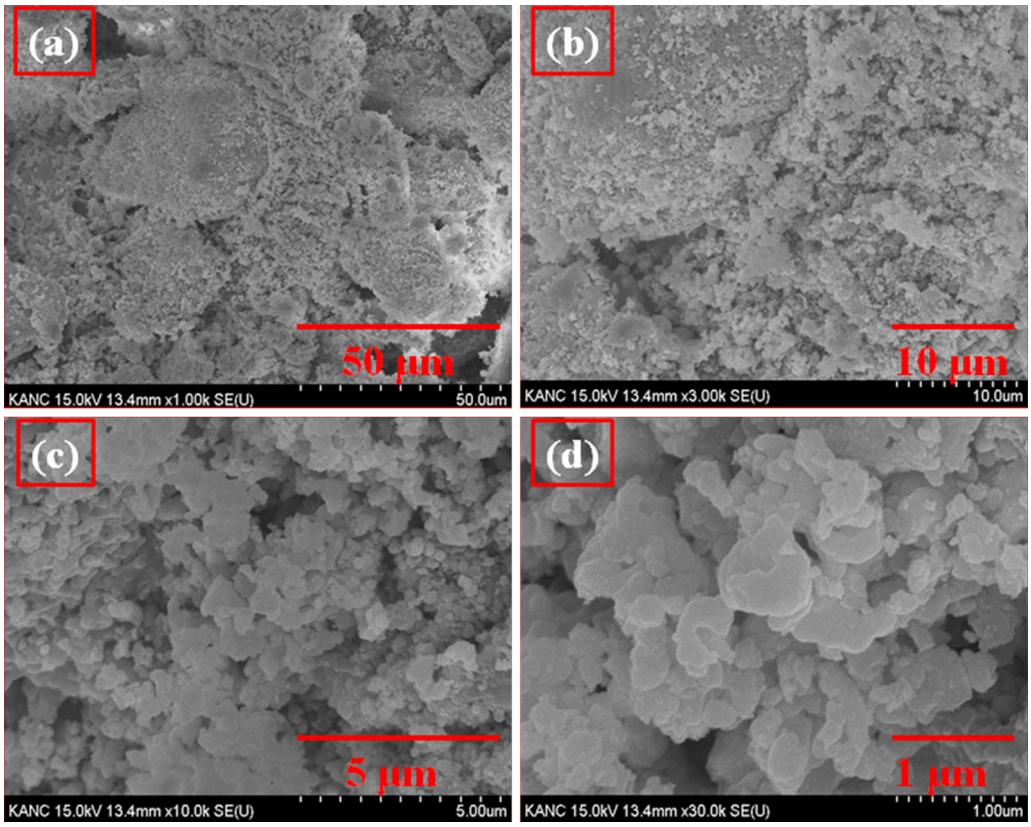


Fig. 2. SEM images (a–d) of $Sr_{1.91}ZnSi_2O_7$:0.09Mn phosphor at different resolutions (1, 3, 10, and 30 K)

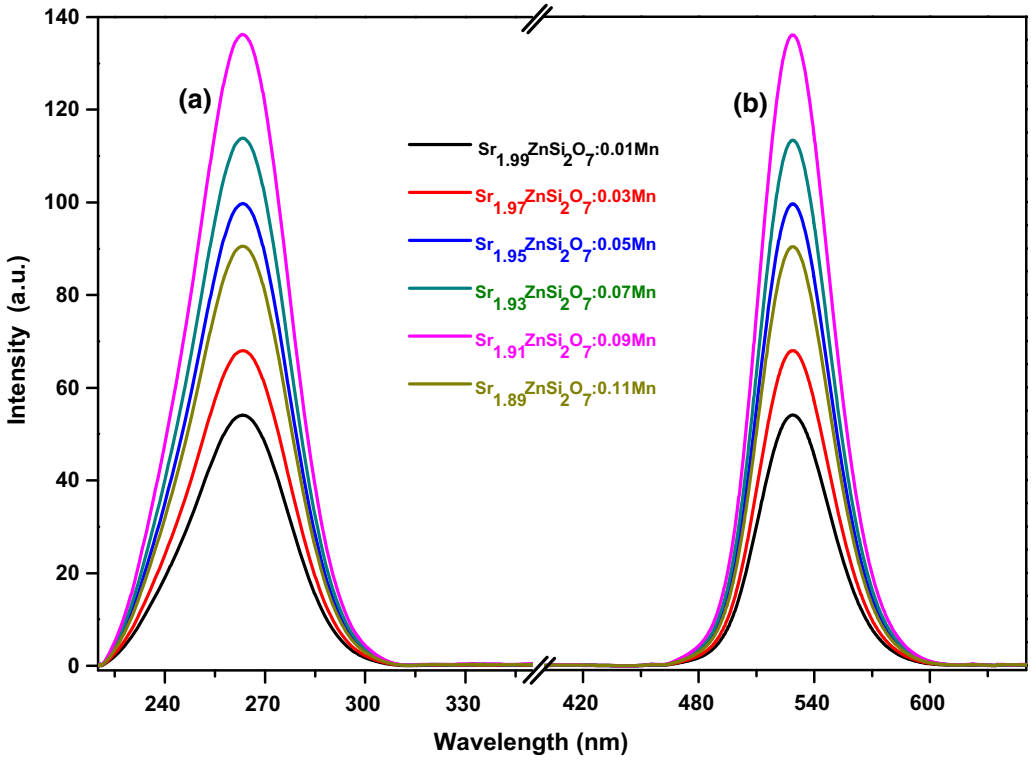


Fig. 3. (a) Excitation spectra of $Sr_{2-x}ZnSi_2O_7$:xMn (λ_{em} = 528 nm) and (b) emission spectra of $Sr_{2-x}ZnSi_2O_7$:xMn (λ_{exc} = 263 nm)

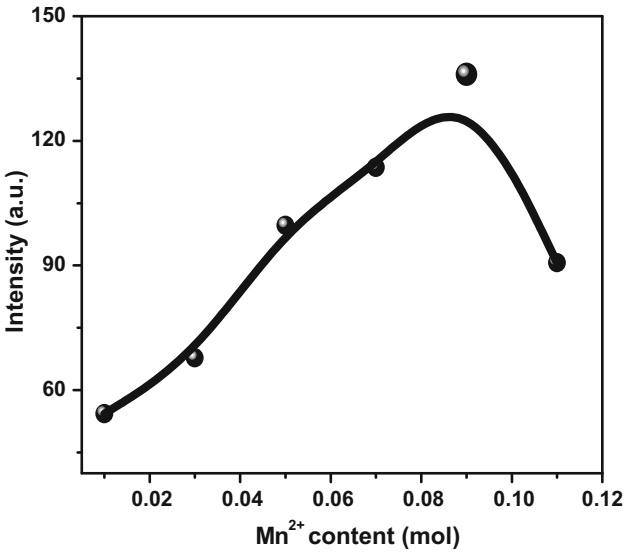


Fig. 4. Variation of PL emission intensity versus Mn^{2+} concentration

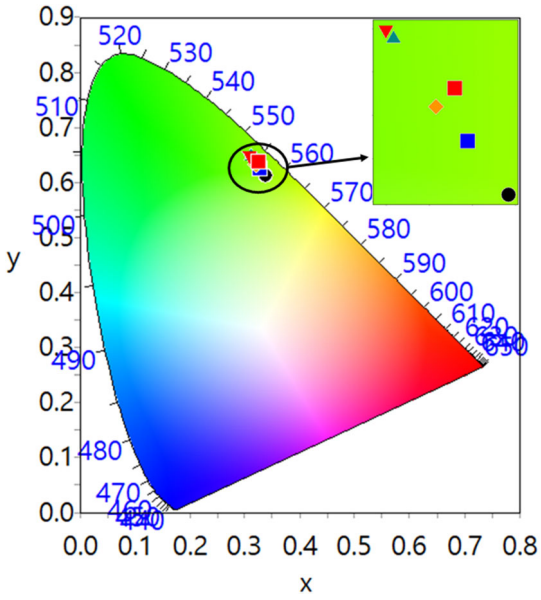


Fig. 5. CIE chromaticity diagram of $Sr_{2-x}ZnSi_2O_7$:xMn (x = 0.01 (●), 0.03 (■), 0.05 (♦), 0.07 (▲), 0.09 (▼) and 0.11 (■)) phosphors. (Inset: magnified image of CIE coordinated specifying the small variation in CIE coordinates with concentration of Mn^{2+} ions in the host).

EPR Analysis

Figure 7 represents the EPR spectra of Mn^{2+} activated $Sr_2ZnSi_2O_7$ phosphors for different concentrations of Mn^{2+} ions (namely 0.01 mol, 0.03 mol, 0.05 mol, 0.07 mol, 0.09 mol and 0.11 mol). The ^{55}Mn nucleus having 5/2 spin interacts with unpaired electrons that leads to fine structure splitting into six lines or sextet hyperfine lines.[26] The EPR spectrum of Mn^{2+} is described by spin-Hamiltonian function as given below:

$$H = \beta BgS + \text{SAI} + \text{ZFS}, \qquad (3)$$

where the first and second term represents the electron spin interaction with the magnetic field and nuclear spin, respectively, and the third term represents zero-field splitting (ZFS) of the sextet ground state. The term ZFS is responsible for the anisotropic part of the EPR spectrum.[27,28] In the present analysis, the EPR resonance signal exhibits six hyperfine lines spread over an average width of 200 mT which are attributed to Mn^{2+} ions.[28] The sextet that originates due to the interaction between Mn^{2+} ions and the defects, slowly broadened with increase in concentration of Mn^{2+} ions in the $Sr_2ZnSi_2O_7$ phosphors.[29] The identical EPR signals signifies the same valence state, identical environment of Mn^{2+} ions in the host thereby indicating spontaneous reduction of Mn^{3+} to Mn^{2+} ions.[29] The EPR spectra exhibit well resolved broad resonance signal centered at $g \approx 2.0$ for all the as prepared samples signify Mn^{2+} ions have occupied deformed cubic sites with tetrahedral symmetry in the present case.[30] Generally, the occupancy of Mn^{2+} ions at tetrahedral sites in a weak crystal field suggests the strong possibility of green emission, while octahedrally coordinated Mn^{2+} ion gives an orange-to-red emission by forming a strong crystal field. Photoluminescence spectra and evaluated CIE coordinates confirms the green emission from the prepared phosphor which is in close agreement with the results obtained from EPR spectra. Moreover, the hyperfine interaction parameter (A) is evaluated from the EPR spectra using the relation mentioned elsewhere.[16] The magnitude of hyperfine interaction parameter depends on the host and electronegativity of the surrounding environment. Moreover,

Table III. CIE color coordinates of $Sr_{2-x}ZnSi_2O_7$:xMn phosphor

Sample composition	*X*-coordinate	*Y*-coordinate
$Sr_{1.99}ZnSi_2O_7$:0.01Mn (●)	0.33847	0.61162
$Sr_{1.97}ZnSi_2O_7$:0.03Mn (■)	0.32889	0.62401
$Sr_{1.95}ZnSi_2O_7$:0.05Mn (♦)	0.32159	0.63199
$Sr_{1.93}ZnSi_2O_7$:0.07Mn (▲)	0.31167	0.64783
$Sr_{1.91}ZnSi_2O_7$:0.09Mn (▼)	0.30992	0.64997
$Sr_{1.89}ZnSi_2O_7$:0.11Mn (■)	0.32595	0.63624

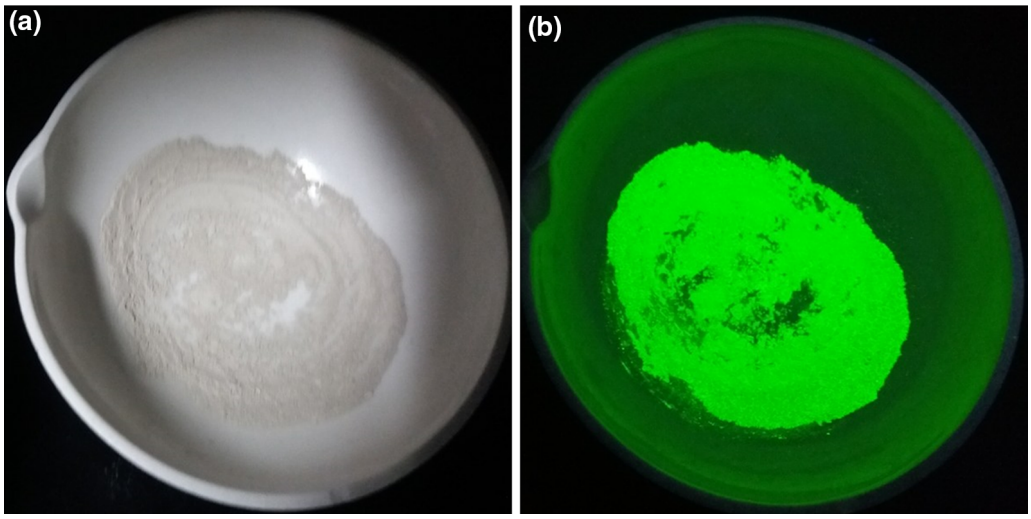


Fig. 6. Digital photograph of $Sr_{1.91}ZnSi_2O_7$:0.09Mn sample (a) under ambient light, and (b) under dark–light with 254 nm excitation

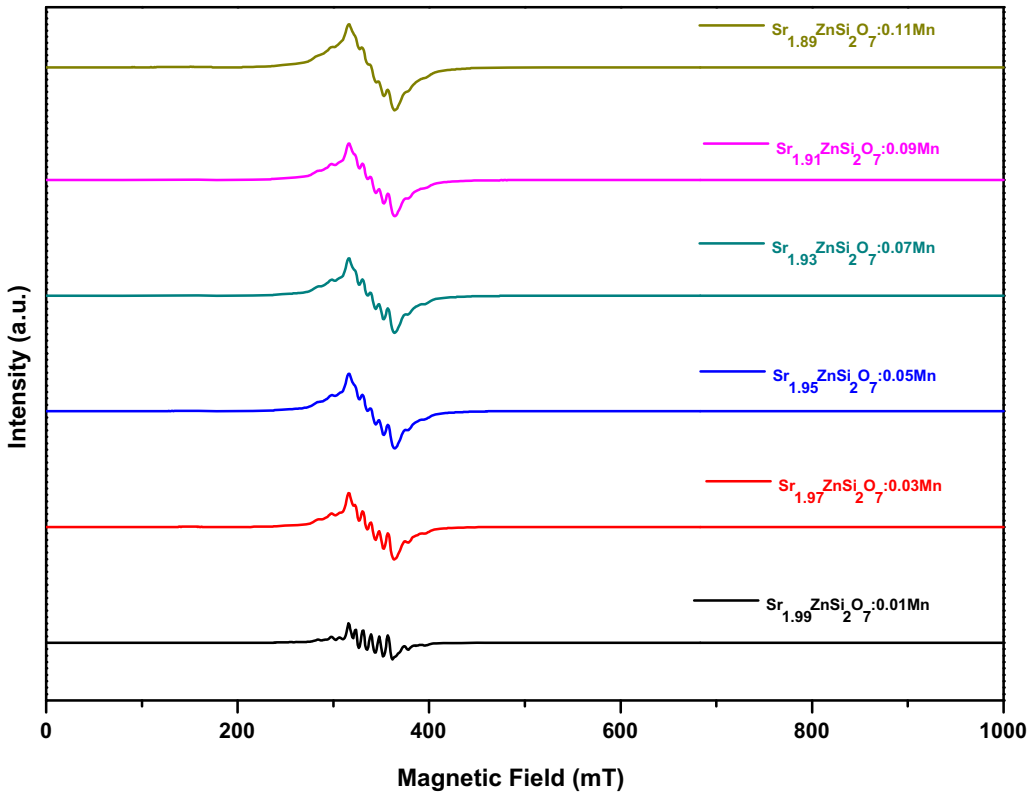


Fig. 7. EPR spectra of $Sr_{2-x}ZnSi_2O_7$:xMn phosphor

it also relates with the nature of bonding between the dopant ion and the ligands bound to the dopant. The higher hyperfine splitting value indicates ionic nature of bonding.[16] The calculated value of A is 88 G indicates ionic bonding between Mn^{2+} ions and anionic ligand. In the present analysis of EPR spectra, the g values and A values are close to those reported by Premkumar et al.[26] The g values for activated $Sr_{2-x}ZnSi_2O_7$:xMn^{2+} (x = 0.01, 0.03, 0.05, 0.07, 0.09 and 0.11) are 2.008, 2.016, 2.021, 2.029, 2.035 and 2.041, respectively. The g values for the Mn^{2+} activated $Sr_2ZnSi_2O_7$ phosphors increases slightly with increase in concentration of Mn^{2+} ions (positive g shift) indicate ionic bonding between Mn^{2+} ions and anionic ligand. The positive g shift observed in the present analysis might be due to electron transfer process through spin–orbit interaction from ligand to Mn^{2+} ion in S-state. Further,

the well resolved hyperfine lines observed in EPR spectra indicate that Mn^{2+} ions are at sufficient distance from each other and occupied distorted tetrahedral symmetry sites.[16,19]

CONCLUSIONS

$Sr_{2-x}ZnSi_2O_7$:xMn (x = 0.01, 0.03, 0.05, 0.07, 0.09 and 0.11) phosphors are synthesized by the sol–gel method by sintering at 1000°C. The XRD patterns confirm the phase formation of $Sr_2ZnSi_2O_7$ phosphors having tetragonal structure. The SEM micrograph shows agglomerated and nanosized particles. The PL spectra of Mn^{2+} ions activated $Sr_2ZnSi_2O_7$ phosphors exhibit intense green emission under UV excitation. The optimized concentration of Mn^{2+} is found to be 0.09 mol and beyond that concentration quenching is achieved. The CIE coordinates further confirm that the $Sr_{1.91}ZnSi_2O_7$: 0.09Mn phosphor exhibits intense green emission under 263 nm excitation. The EPR spectra reveal resonance signals at $g \approx 2.0$, where Mn^{2+} ions have occupied tetrahedral symmetry sites. With increase in Mn^{2+} ion concentration, there is a slight increase in g values that signifies ionic bonding between Mn^{2+} and O^{2-} ions. All these results prompted us to conclude that Mn^{2+} activated $Sr_2ZnSi_2O_7$ phosphor can serve as a green emitting phosphor for its utility in display devices.

ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1A1B03030003).

REFERENCES

1. Y.D. Jiang, F. Zhang, and C.J. Summers, *Appl. Phys. Lett.* 74, 1677 (1999).
2. Y. Hu, W. Zhuang, H. Ye, D. Wang, S. Zhang, and X. Huang, *J. Alloys Compd.* 390, 226 (2005).
3. T. Hase, T. Kano, E. Nakazawa, and H. Yamamoto, *Adv. Electron. Electron Phys.* 79, 271–373 (1990).
4. W. Zhao, G. Ristic, and J.A. Rowlands, *Med. Phys.* 31, 2375–2726 (2004).
5. C.H. Kim, I.E. Kwon, C.H. Park, Y.J. Hwang, H.S. Bae, B.Y. Yu, C.H. Pyun, and G.Y. Hong, *J. Alloys Compd.* 311, 33–39 (2000).
6. L. Bindi and P. Bonazzi, *Phys. Chem. Miner.* 32, 89–96 (2005).
7. L.H. Merwin, A. Sebald, and F. Seifert, *Phys. Chem. Miner.* 16, 752–756 (1989).
8. S. Yao, L. Xue, and Y. Yan, *J. Electroceram.* 26, 112–115 (2011).
9. T. Joseph, S. Uma, J. Philip, and M.T. Sebastian, *J. Mater. Sci. Mater. Electron.* 22, 1000–1009 (2011).
10. T. Joseph, S. Uma, J. Philip, and M.T. Sebastian, *J. Mater. Sci. Mater. Electron.* 23, 1243–1254 (2012).
11. Y. Hao and Y.H. Wang, *Mater. Res. Bull.* 42, 2219–2223 (2007).
12. X. Wang, Z.Y. He, D. Jia, W. Strek, R. Dariusz, D. Hreniak, and W.M. Yen, *Microelectron. J.* 36, 546–548 (2005).
13. Y. Zhang, P. Ran, L.I. Chengyu, Z. Chunyu, and S.U. Qiang, *J. Rare Earths* 28, 705–708 (2010).
14. K. Ravindranadh, R.V.S.S.N. Ravikumar, and M.C. Rao, *AIP Conf. Proc.* 1728, 020079 (2016).
15. Y. Chen, Y. Li, J. Wang, M. Wu, and C. Wang, *J. Phys. Chem. C* 118, 12494–12499 (2014).
16. V. Singh, R.P.S. Chakradhar, J.L. Rao, and D.K. Kim, *J. Lumin.* 128, 1474–1478 (2008).
17. B. Tripathi, F. Singh, D.K. Avasthi, D. Das, and Y.K. Vijay, *Physica B* 400, 70–76 (2007).
18. K. Li, H. Wang, X. Liu, W. Wang, and Z. Fu, *J. Eur. Ceram. Soc.* 37, 4229–4233 (2017).
19. V. Singh, R.P.S. Chakradhar, J.L. Rao, and D.K. Kim, *Mater. Chem. Phys.* 110, 43–51 (2008).
20. X. Li, X. Tang, Z. Wang, Z. Zou, J. Zhang, Z. Ci, and Y. Wang, *J. Alloys Compd.* 721, 512–519 (2017).
21. V. Singh, G. Sivaramaiah, J.L. Rao, and S.H. Kim, *J. Lumin.* 157, 74–81 (2015).
22. N. Singh, V. Singh, G. Sivaramaiah, J.L. Rao, P.K. Singh, M.S. Pathak, S.J. Dhoble, and M. Mohapatra, *J. Lumin.* 178, 479–486 (2016).
23. M. Ardit, G. Cruciani, and M. Dondi, *Z. Kristallogr.* 225, 298 (2010).
24. H. Yan, *J. Wuhan Univ. Technol. Mater. Sci. Ed.* 30, 97–99 (2015).
25. Z. Zhang, C. Ma, R. Gautier, M.S. Molokeev, Q. Liu, and Z. Xia, *Adv. Funct. Mater.* 28, 1804150 (2018).
26. H.B. Premkumar, D.V. Sunitha, H. Nagabhushana, S.C. Sharma, B.M. Nagabhushana, C. Shivakumara, J.L. Rao, and R.P.S. Chakradhar, *J. Lumin.* 132, 2409–2415 (2012).
27. S.V. Nistor, L.C. Nistor, M. Stefan, C.D. Mateescu, R. Birjega, N. Solovieva, and M. Nikl, *Superlattices Microstruct.* 46, 306–311 (2009).
28. V. Singh, R.P.S. Chakradhar, J.L. Rao, and D.-K. Kim, *J. Lumin.* 129, 755–759 (2009).
29. Y. Wei, X. Han, E. Song, and Q. Zhang, *Mater. Res. Bull.* 113, 90–96 (2019).
30. M.R. Yadav, P.R. Krishna, B.T. Rao, B.R.V. Rao, and G.V. Lakshmikanth, *Mater. Today Proc.* 5, 25807–25814 (2018).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Life Cycle Reliability and Safety Engineering
https://doi.org/10.1007/s41872-020-00110-7

ORIGINAL RESEARCH

Estimating maintenance budget using Monte Carlo simulation

Atul Kumar Srivastava[1] · Girish Kumar[1] · Piyush Gupta[2]

Received: 16 November 2019 / Accepted: 8 January 2020
© Society for Reliability and Safety (SRESA) 2020

Abstract
Asset maintenance is driven by allocation of yearly maintenance budget (YMB). Lack of scientific methods for its evaluation has led to allocations that are subjective and erroneous causing wastage or lack of resources. The objective of the work is to estimate yearly budget for plant maintenance. In this work, the constraint of subjective quantification of budget influencing parameters proposed in the literature is removed. Suggested methodology is more realistic in a real world context and results into more rational estimation of maintenance budget. Graph theory and matrix approach model along with Monte Carlo simulations are deployed to evaluate maintenance budget by assigning random values to the budget parameters within specified range. MatLab code is deployed for its implementation. The budget influencing parameters are simulated for pre-defined maintenance levels. Inferential statistics is applied to find the appropriate distribution and evaluate the interval estimation of maintenance budget. The maintenance budget so estimated is validated by comparing it with world-class and best practices values reported in the literature. The Monte Carlo simulation results for different maintenance levels are presented with mean, variance and interval estimation of maintenance budget. The methodology suggested in this work will be a useful aid to maintenance managers in assigning YMB to their plant systems and provide indications for maintenance system corrections. This will avoid using historic data or subjective expert judgments.

Keywords Asset management · Yearly maintenance budget · Monte Carlo simulation · Graph theory · Decision-making · Maintenance management

1 Introduction

Maintenance signifies upkeep of the assets through the use of proactive, reactive and other maintenance strategies (Swanson 2001). It also ensures quick equipment turnaround to minimize downtime and enhance the availability of physical plant assets. Maintenance is a challenging task, many a times needing out of the box thinking and deployment of optimization algorithms using software modules, for example: inventory management, manpower allocation, maintenance scheduling, etc. Manufacturing firms have realized the importance of this profit enhancing activity, as plant and equipment are becoming increasingly complex and timely maintenance interventions can prevent catastrophic failures.

Maintenance budget (MB) is a driver for maintenance work. It signifies the availability of optimum financial resources to perform the maintenance activities. Budget for maintenance is needed for arranging maintenance workforce, spares, consumables, tools, special equipment, etc. Lack of appropriate maintenance budget can throttle maintenance works. On one hand, availability of excess budget can lead to wasteful expenditure; while on the other hand, lack of financial resources can create situations, wherein there are production outages and unsafe plant practices. Therefore, estimation of appropriate maintenance budget is essential for organizations to ensure that they maintain their competitiveness. The existing methodology for estimating maintenance budget is one of setting MB as some percentage of the asset cost; using historical costs; and using expert judgment. All of these methods are non-scientific and might lead

✉ Girish Kumar
girish.kumar154@gmail.com

Atul Kumar Srivastava
atchan1@rediffmail.com

Piyush Gupta
piyushiit1980@gmail.com

[1] Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, New Delhi, Delhi 110042, India

[2] Retired Engineer - G, Inter University Accelerator Centre, New Delhi 110067, India

Published online: 22 January 2020

to wastage of resources. The reasons for wide variations in budget are many, such as variety of plant assets, safety levels, age and complexities of the plant, operating conditions, etc. Maintenance budget is generally evaluated as % of asset replacement value (ARV) (Gulati and Smith 2009; Gupta et al. 2014). ARV is the current replacement cost of asset including the cost of removing the old assets and the cost of installing new one. High YMB as a % of ARV that leads to large maintenance expenditures do compel the organizations to revisit their maintenance management systems and strike a balance between the need to maintain vis-à-vis asset replacement.

Yuzvovich et al. (2018) discussed three methods, namely: statistical monitoring, analysis and measurement, and observational generalization and logical inference for system of budgeting at industrial enterprises. (O'Leary et al. 2018) suggested a partial budgeting tool to assess investments in Precision dairy technologies. (Gupta et al. 2014) proposed a methodology based on graph theory and matrix approach for the estimation of YMB of a plant system under its contextual conditions. They considered eleven budget influencing parameters. These parameters had numerous sub-parameters too. This work is an extension of the previous work. In this, an attempt has been made to allocate a range to each budget influencing parameters as against a point value allotted to each parameter in the earlier work. Monte Carlo simulation is run wherein each budget parameter can pick any random value within the specified ranges to arrive at YMB corresponding to the random data set of the parameters. Such a scenario is more realistic in a dynamic and probabilistic context as the YMB parameter may not have fixed values spread over the entire budgeting year.

The next Sect. 2 discusses some of the relevant papers available in literature.

2 Literature review

In this section, literature review is carried out on estimation of maintenance budget and Monte Carlo simulation technique that is used to evaluate the YMB.

2.1 Maintenance budget

Numerous studies have been reported in the literature concerning the maintenance costs and maintenance budgeting. Some of these are discussed in this sub-section.

Maintenance costs account for 15–30% of today's manufacturing costs (Chan et al. 2005). Sarker and Faiz (2016) showed that total cost of maintenance is dependent on number of age groups and age threshold sets of components. Maintenance cost in mining industry is reported to be 40–50% of the total operating cost (Campbell 1995). Manufacturing industry in UK reports to have spent 12–23% of the total factory operating costs on maintenance. Dekker (1996) showed that over the years, significant additional capital has been deployed in production equipment to automate and mechanize the production processes. As a result, the manpower on production has reduced, and there is a considerable increase in employees working in the maintenance area. Therefore, maintenance spending as a % of the total operational costs is gradually increasing. Past experiences have shown that optimal scheduling of preventive maintenance interval outages maintenance cost of a plant, while increasing reliability of system. Loganathan and Gandhi (2016) suggested particle swarm optimization technique for minimization of the maintenance cost of manufacturing system under reliability constraints. Alsyouf (2009) investigated the maintenance practices that are used in Swedish industry by conducting a survey within firms that have at least 100 employees, and their study showed that the role of maintenance is not yet given its due importance. They found that 0–20% of maintenance budget was spent on education and training with a mean of about 4%. The research used exploratory factor analysis with varimax rotation and identified 13 variables/aspects and seven main maintenance practices that the respondents considered important while selecting a maintenance policy. Finally, it was suggested that the greater adoption of maintenance concepts such as Total Productive Maintenance (TPM) and Reliability-Centered Maintenance (RCM) must be done to enhance operational efficiency. Study of literature revealed that there is a dearth of data related to maintenance costs and budgeting and the figure varies widely across the equipment and plant domain, and is primarily context dependent, such as: operating conditions; criticality; equipment age; workmen skills; efficacy of maintenance planning; the economic trade-off between reliability and safety versus the maintenance costs that are incurred, etc. (Gupta et al. 2014). However, (Gulati and Smith 2009) did indicate that the world-class organizations spend about 2.5–3.5% of asset replacement value (ARV) on maintenance, and the benchmark lies between 3 and 9%, if best maintenance practices are followed.

Maintenance cost is a key to competitiveness of any plant. Benchmark and optimization models must be established to achieve it. Several researchers have attempted to evaluate the YMB and devise models and algorithms to figure out its value. It is well recognized that appropriate maintenance policy significantly reduces the maintenance cost/budget. Topal and Ramazan (2010) showed that optimal scheduling of mining equipment in the industry using mixed integer programming reduces the maintenance cost by 16% in 10 years. Bahr and Lennerts (2010) developed a Practical Adaptive Budgeting of maintenance to calculate maintenance budget, which attempts to eliminate the shortcomings of the existing methods. They found that maintenance cost peaks in

30 years of asset life. Ottoman et al. (1999) estimated the maintenance budget on the basis of the plant's initial cost or replacement value; quantifiable attributes of the plant; life-cycle costs of systems and components, and assessment of current or projected physical condition of the asset and compared the eighteen maintenance and repair budgeting models. It is also experienced that maintenance budget largely depends on environment, such as, harsh climatic condition, lack of suitable and sufficient infrastructure, etc. Kayrbekova et al. (2011) considered these aspects and showed that sub-zero temperatures have a great influence on reliability and maintainability performance and, therefore, the maintenance cost. Komonen (2002) proposed a cost model of industrial maintenance for profitability analysis and benchmarking. He identified the various performance indicators and developed a hierarchical system of key Figure/indicators. The cost model was tested with data collected from 436 production units and causalities were established between variables. On the basis of this, he developed a benchmarking tool for expected maintenance cost of the production unit and the impact of maintenance cost on the competitiveness of a plant. Dordevic et al. (2012) established an optimization model for maintenance budget, which was based on maintenance plan. The model minimizes the maintenance budget and achieves the required level of reliability. Ferreira et al. (2013) presented a non-linear programming formulation for the maintenance budgeting problem for a multi-component system. The result of this formulation minimizes the expected value of the sum of stochastic and deterministic costs. Nahas et al. (2008) analyzed the optimal sequence of preventive maintenance actions, which minimizes the maintenance cost with desired level of system reliability for a multi-state system. Pham and Wang (1996) in their research discussed the optimal maintenance policies that aim to provide optimum system reliability or availability and safety performance at the lowest possible maintenance cost. If the YMB is more than 20% of ARV, then maintenance interventions are not justified and maintenance manager must consider replacing the plant by a new one.

Further, (Gupta et al.2014) developed a methodology to calculate the YMB for a plant system as a percentage of its asset replacement value (ARV). They identified the various parameters influencing the budget and modeled these in terms of a Plant Maintenance Budget Digraph. Their technique facilitated the maintenance managers to evaluate the YMB, which is based on plant-specific values of parameters that vary from one plant to another and are contextual in nature. The model was validated using several case studies. Additional validation of this model was also done by (Gupta 2017) in evaluating the maintenance cost for a particle accelerator system. In another similar work, (Gupta and Gupta 2017) evaluated the YMB for an air conditioning plant under its contextual conditions. They determined various parameters affecting the maintenance budget and using the incidence matrix, they evaluated the YMB. The results showed that the model proposed by the authors in (Gupta et al.2014) could evaluate the maintenance budget, without resorting to expert judgment. The methodology was capable of identifying the weaknesses in the maintenance system, but was strongly constrained by the salvage value of the old asset. Moreover, it fell short of evaluating the maintenance budget for maintenance actions arising out of failures that are stochastic in nature. Though the model developed by (Gupta et al. 2014) has stood the validations well vis-à-vis the values reported in the literature, the model needs to evaluate mean, distribution and variance of the YMB based on random allocation of values to the parameters, which actually happens in practice. Monte Carlo simulation (MCS) technique has been used to arrive at these. The literature review on the technique of MCS deployed in this work is discussed in the next sub-Sect. 2.2.

2.2 Monte Carlo simulation (MCS) technique

The MCS enables one to evaluate system parameters that cannot be easily evaluated using analytical models. MCS uses the computer to generate uniformly distributed pseudo-random numbers and convert them into real-world responses using inverse function under existing mathematical models and assumed probability distribution.

Few applications that have used MCS in the domain of maintenance are discussed here. Condition-based maintenance policies are of central importance in the continuous monitoring of the condition of modern plants and industries. It enables maintenance managers to make dynamic maintenance decisions based on observed condition of the system. Marseguerra et al. (2002) determined optimal degradation level beyond which preventive maintenance must be carried out. They use continuously monitored multi-component system for modeling and solved the problem by means of Genetic algorithm coupled with Monte Carlo simulation. Cantoni et al. (2000) used MCS coupled with a Genetic Algorithm (GA) technique for optimal plant design under economic and safety constraints. Borgonovo et al. (2000) developed a MCS approach to plant availability by modeling maintenance, aging and obsolescence. Marseguerra and Zio (2000) established a stochastic model of plant operation for optimizing maintenance and repair policies using a combination of GA and MCS. Hall and Strutt (2003) in their research evaluated characteristic life and shape factor using MCS. Marquez et al. (2004) used MCS for reliability and availability assessment of complex systems and presented a generic approach using continuous time MCS modeling. Ramakrishnann (2016) suggested the use of MCS for estimating the unavailability and mean time to failure of a fast

reactor. Cadini and Gioletta (2016) proposed Bayesian MCS for estimating small failure probabilities.

Individuals, companies as well as society are becoming increasingly dependent on complex systems. Failure of these often causes a major downtime with potentially serious consequences. Crespo and Iung (2007) developed a MCS-based model to assess the availability and reliability of a system when various factors like complexity of system, range of failure modes, environment, and sustainability affect the system behavior. MCS has become all pervasive since its introduction, but for its computational cost. To reduce that computational expense, the sampling efficiency and convergence for MCS were investigated by (Janssen 2013).

Study of literature showed that the evaluation of YMB till now is usually based on contextual conditions with specific point values assigned to budget influencing parameters. The present work removes this constraint and attempts to assign values to the parameters within certain broad ranges. It tries to estimate the YMB along with its distribution, mean and variance, which corresponds to certain chosen maintenance levels. MCS are run on the already developed model based on graph theory and matrix approach (GTMA). It evaluates the maintenance budget of any plant working in any specific and broad maintenance levels. This will be a useful aid to maintenance managers in assigning YMB to their plant systems.

The paper is divided into seven sections. The next Sect. 3 briefly discusses the parameters influencing the YMB. Section 4 discusses the methodology, which includes the model description, its algorithm and the MATLAB code developed for the purpose. Section 5 estimates the YMB under different maintenance scenarios. Section 6 evaluates the distribution of YMB function and interval estimates for mean and variance. Section 7 discusses the practical implications and finally the last Sect. 8 concludes the paper.

3 Parameters influencing YMB

The YMB is considered a system that is dependent on several parameters as shown in Fig. 1. These are briefly discussed below.

Plant availability is the probability that a system is performing its required function at a given point in time when operated under stated operating conditions. Opportunistic maintenance plays a vital role in increasing plant availability by utilizing plant shutdown window to replace or upgrade the partially degraded subsystems/units (Kumar et al. 2014). Plant availability greater than 90% signifies a world-class operating standard for the plant. Industrial plant operations are intrinsically hazardous. Therefore, hazardous plants will demand a stringent maintenance

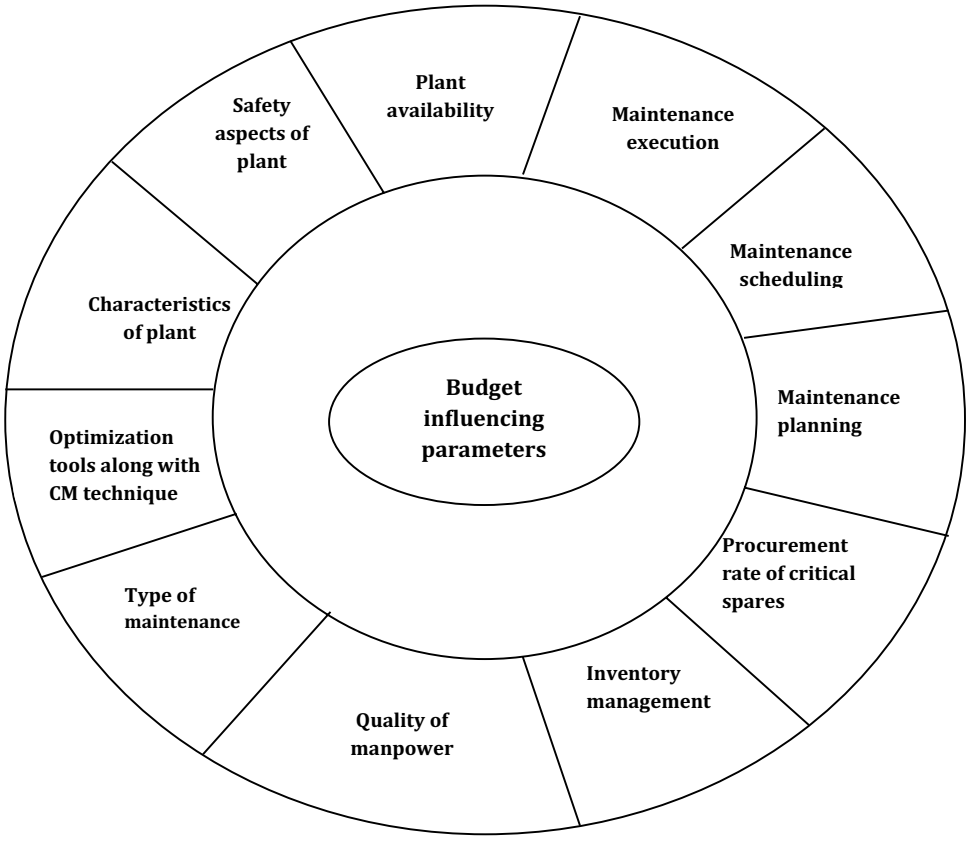


Fig. 1 Budget influencing parameters

work to overcome risk of accidents (Loganathan et al. 2018a, b). Higher maintenance budget is, therefore, required to maintain the proper safety of a plant. Physical environment, complexity of plant, redundancy in plant system, level of technology leading to operational automation, service duty of plant, plant layout, physical location of plant, etc. are some of the parameters that characterize a plant maintenance system. These influence the maintenance budget.

It is well established that the maintenance expenditure can be optimized using mathematical models that includes condition monitoring techniques. Plant availability is enhanced by carrying out maintenance at optimum interval of preventive and condition-based maintenance (Kumar and Varghese 2018; Kumar et al. 2018a). These help in evaluating the maintenance strategies to be deployed including the scheduling of maintenance tasks.

Type of maintenance work is also a relevant variable that influences the maintenance budget. Proactive maintenance as against resorting to reactive maintenance ensures higher availability of plant. It is recognized that proactive maintenance reduces the maintenance cost by 15–30% and downtime by 20–40%, while increasing the production output by 15–25%. It is further recommended that more than 80% of maintenance work must be proactive (Gulati and Smith 2009).

Kumar and Gandhi (2011) reported that majority of equipment in any plant undergo breakdowns due to human errors. Therefore, skilled manpower is essential to the maintenance activities and it strongly influences YMB.

It is well established that an appropriate inventory management system and supplier relationship reduces the waiting time for spares during execution of the maintenance activities and is, therefore, significant in reducing maintenance costs. (Gulati and Smith 2009) reported that maintenance personnel usually waste 30% of their time in locating spare parts. Therefore, effective material management software must be used by the organizations for their stocking decisions to prevent unnecessary delay in maintenance work. Condition-based maintenance decreases requirement of in-stock inventory as the procurement of parts can be triggered by the identification of a potential failure, (Louit et al. 2011).

Maintenance planning based on the work content, methodology, technical documentation, start date, required by date, man/material resources, safety requirements and special requirements, etc. minimizes the probability of system failure and its consequences related to safety, failure train and environment. It helps management in taking decisions concerning investment in maintenance. This will result in better plant utilization. It is well known that simultaneous planning of production and maintenance does result in reducing the maintenance budget. Scheduling intensity, scheduling quality and schedule realization rate are the key indicators to quantify the maintenance scheduling variable (Muchiri et al. 2011).

Execution of maintenance tasks is done after planning and scheduling of the tasks as per operational requirements. The YMB is also dependent on maintenance execution variable, which considers the values of scheduling compliance, percentage of re-work, ratio of completed to received tasks, manpower efficiency and number of overdue tasks (Muchiri et al. 2011).

This section described the parameters influencing YMB and the next section covers the methodology for evaluating the YMB including model description and algorithm.

4 Methodology

This section briefly describes the YMB model proposed by Gupta et al. (2014). This section also describes the computer algorithm developed for running MCS on the developed model by removing the constraint of contextual and point values of budget influencing parameters.

The YMB system was modeled by identifying the parameters influencing it. The values of the parameters were assigned in percentage based on tables developed for their quantification by the authors. A higher value of the parameters implies reduction in the maintenance budget. The value 100 represents the ideal condition of parameters.

In this work, the constraint on parameter quantification is removed to avoid subjectivity involved in the quantification methodology proposed in the earlier work. Monte Carlo simulations are run by assigning any random value to the budget parameters within the chosen ranges. Maintenance budget is estimated for 'world-class practices' and six other levels namely L_1, L_2, L_3, L_4, L_5 and L_6 in ascending order of maintenance expenditure. Best practices are approaches that have been applied by organizations who are pioneers in their industry. It is a methodology utilized by the most competitive, and efficient organization. Such companies are high-quality manufactures with very competitive prices typically lowest in their industry. Executing best maintenance and reliability practices will enable an organization to: achieve better efficiency with the same set of asset, reduce maintenance and overall cost per unit output, enhance performance, increase competitiveness and reduce the need to replace the capital. For the organization that meets best practices benchmark, the maintenance budget as a percentage of asset replacement value is 2–9%. World-class practices are the best of best practices and the maintenance budget varies from 2 to 3.5% for the organization adopting world-class practices. It has already been addressed that a plant's maintenance budget relies on the conditions of maintenance parameters. As the value of maintenance parameters reaches towards 100, the plant hits to a world-class level and

shifts to a level of best practices, good practices and worst practices when quality of parameters moves away from the ideal condition. In ideal condition, the value of maintenance parameter is 100. In the present study, the categorization of maintenance level is based on the value (in range) of maintenance parameters listed in Table 1. L_1 is the level with a high maintenance parameter range and L_6 has a low maintenance parameter range.

4.1 Model description

The proposed work is the extension of the budget model suggested by (Gupta et al. 2014) by removing the constraint of assigning point values to budget parameters.

The steps followed in formulating the model are briefly discussed as below:

- As a first step, the budget influencing parameters were identified.
- Next, the budget matrix was developed considering the identified parameters and their degree of mutual influences.
- The budget matrix of size n x n (n = number of budget influencing parameters) had diagonal elements, which were logarithm of the percentage value of budget parameters and the values of the off-diagonal elements were the degree of influences among the parameters. The degree of influence was quantified on a scale of 0–3, with 0 signifying no influence; while 1, 2, and 3 signifying a weak, medium and strong influence.
- Finally, the Permanent Function (Jurkat and Ryser 1966) for the matrix was obtained and was named as maintenance budget function (MBF). Use of Permanent Function ensured that no information was lost.
- The YMB is then evaluated from the Eqs. (1) and (2).

$$\text{Maintenance Budget Index(MBI)} = \frac{\text{MBF}_{\text{Ideal}}}{\text{MBF}_{\text{actual}}}, \tag{1}$$

$$\text{YMB} = (\text{MBI} \times 100 - 100)\% \text{ of ARV}. \tag{2}$$

Gupta et al. (2014) in their model evaluated the values of diagonal elements by quantifying them using tables developed for the purpose. In this work, the YMB, however, is evaluated without resorting to the quantification tables and by assigning random values to the budget parameters within specified ranges and deploying MCS simulations to arrive at steady-state means of YMB. The algorithm for the simulation is described in the next sub-sections. A MATLAB code is developed for its implementation.

4.2 Algorithm

An algorithm is developed for YMB estimation based on MCS and various steps are detailed below.

Step (i)	$X_1, X_2, X_3 \ldots X_n$ are the n budget influencing parameters. These parameters are diagonal elements of the matrix and input to the model
Step (ii)	Off-diagonal elements of matrix for each variable are calculated depending upon the intensity of relation with other parameters
Step (iii)	Allocate the value log100% = 2 to all budget influencing parameters under ideal condition
Step (iv)	Generate the ideal maintenance budget matrix of size $n \times n$
Step (v)	Evaluate Permanent of matrix under ideal condition

Table 1 Maintenance budget parameter ranges for various maintenance scenarios

S. No.	Maintenance budget influencing parameters	World-class practices (typical range)	World-class practices (minimum threshold)	Maintenance budget parameters ranges					
		L_{WC}	L_{WCmin}	L_1	L_2	L_3	L_4	L_5	L_6
1.	Plant availability	90–100	90	90–100	80–100	70–100	80–90	70–80	60–70
2.	Safety aspects of plant	100	100	90–100	80–100	70–100	80–90	70–80	60–70
3.	Characteristics of plant	93–100	93	90–100	80–100	70–100	80–90	70–80	60–70
4.	Optimization tools along with condition monitoring technique	100	100	90–100	80–100	70–100	80–90	70–80	60–70
5.	Type of maintenance	80–100	80	90–100	80–100	70–100	80–90	70–80	60–70
6.	Quality of manpower	95–100	95	90–100	80–100	70–100	80–90	70–80	60–70
7.	Inventory management	98–100	98	90–100	80–100	70–100	80–90	70–80	60–70
8.	Procurement rate of critical spares	99–100	99	90–100	80–100	70–100	80–90	70–80	60–70
9.	Maintenance planning	95–100	95	90–100	80–100	70–100	80–90	70–80	60–70
10.	Maintenance scheduling	95–100	95	90–100	80–100	70–100	80–90	70–80	60–70
11.	Maintenance execution	90–100	90	90–100	80–100	70–100	80–90	70–80	60–70

Table 2 Mean and variance of YMB as % of ARV for different plant category

	Range of maintenance parameters for various maintenance levels	Mean of YMB as % of ARV	Variance of YMB as % of ARV	Standard deviation as % of ARV
World-class practices	Range of parameters as per Table 1	1.71	0.28	0.53
World-class practices corresponding to point values of budget parameters at minimum threshold	Values of parameters as per Table 1	3.5	–	
Maintenance levels	L_1	3.20	0.70	0.84
	L_2	6.82	3.40	1.84
	L_3	7.50	10.76	3.28
	L_4	10.52	1.04	1.02
	L_5	19.49	1.63	1.28
	L_6	30.83	2.74	1.65

Step (vi)	Generate the random numbers between the chosen range of parameters and allocate the random value to each variable
Step (vii)	Generate the maintenance budget matrix for a chosen maintenance level corresponding to levels detailed in Table 1
Step viii)	Evaluate Permanent of matrix arrived at in step (vii) above
Step (ix)	Calculate the MBI from Eq. (1) for the chosen maintenance level
Step (x)	Evaluate the YMB from Eq. (2) for the chosen maintenance level
Step (xi)	Repeat the step (vi) through step (*x*) for N number of cycles till the YMB value converges.
Step (xii)	Calculate mean and variance of YMB after each run
Step (xiii)	Simulate the mean for N numbers of cycles using MCS
Step (xiv)	Perform statistical analysis on model output

Based on the algorithm suggested in this sub-section, a MATLAB code is developed for evaluation of YMB.

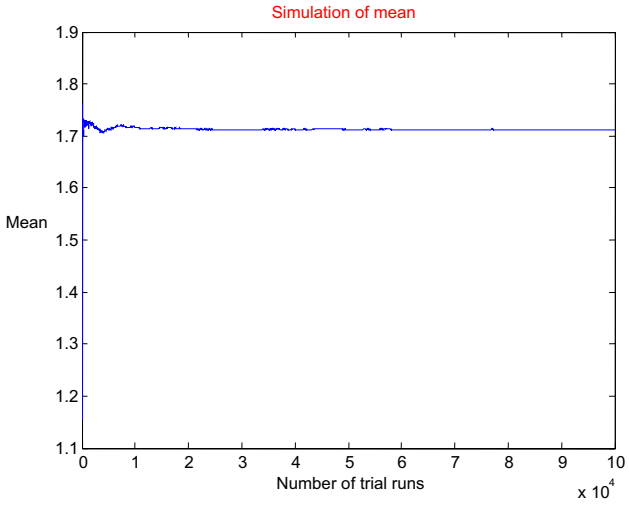


Fig. 2 Mean of YMB versus number of simulations for world-class practices level of maintenance

5 Estimation of yearly maintenance budget

Following the discussion in Sects. 3 and 4, the mean and the variance of YMB as a percentage of asset replacement value (ARV) are evaluated in this section. Mean and variance of YMB are evaluated for different maintenance levels operative in the plant systems. The classifications of different maintenance levels based on the input data range of parameters are discussed in Sect. 3 and given in Table 1. To implement the suggested methodology, the MATLAB code was deployed. The simulations were run for 10^5 trials for convergence of the results. These are tabulated in Table 2. Result for the maintenance level 'world-class practices' and six other levels, namely L_1, L_2, L_3, L_4, L_5 and L_6 are shown graphically also in Figs. 2, 3, 4, 5, 6, 7 and 8.

A closer look at the results shown in Table 2 indicates that mean of YMB as percetage of ARV is 1.71 for

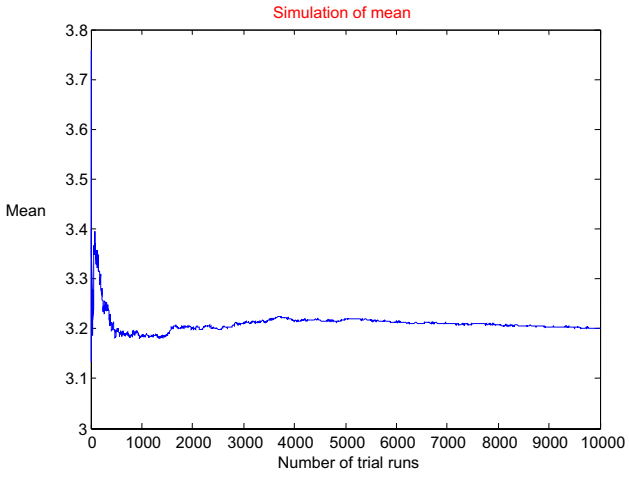


Fig. 3 Mean of YMB versus number of simulations for L_1 level of maintenance

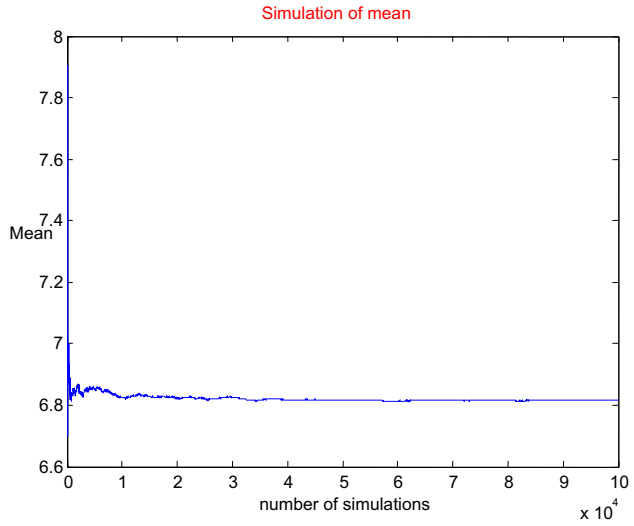


Fig. 4 Mean of YMB versus number of simulations for L_2 level of maintenance

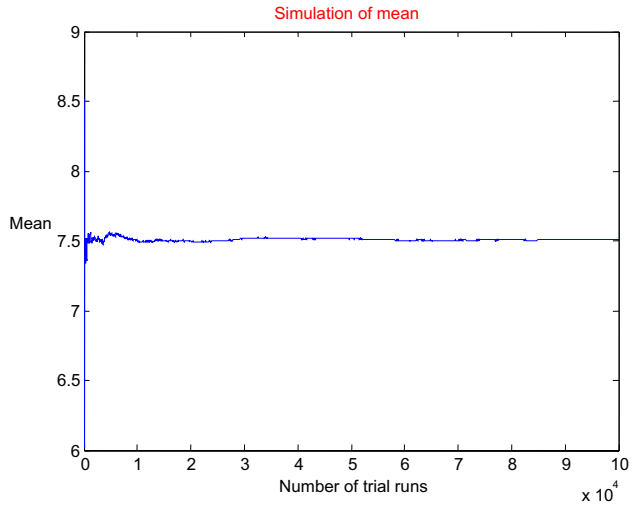


Fig. 5 Mean of YMB versus number of simulations for L_3 level of maintenance

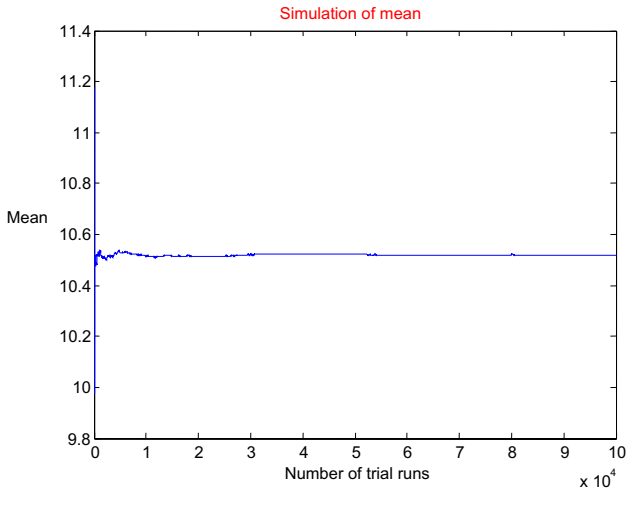


Fig. 6 Mean of YMB versus number of simulations for L_4 level of maintenance

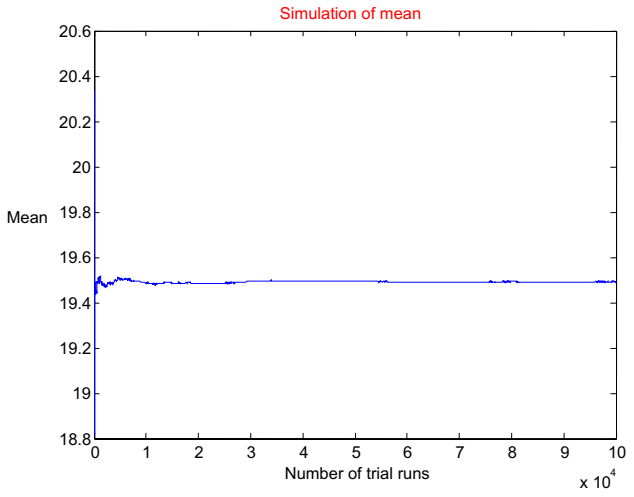


Fig. 7 Mean of YMB versus number of simulations for L_5 level of maintenance

maintenance budget parameters having world-class values as per the ranges specified in the literature (Gupta et al. 2014). This is in contradiction to the range of 2–3.5% as reported in the literature (Gulati and Smith 2009). The authors are not aware of the mechanism that Gulati and Smith had used to indicate the budget influencing parameters range of world-class practices. Also, authors apprehend that Gulati and Smith may have used point values of parameters on the lower side of world-class bound or else they may have arrived at these figures through an extensive survey of the industries following world-class practices. Further, the YMB for point values of budget influencing parameters at their minimum threshold of world-class values is evaluated as 3.5% of ARV, which matches with the upper bound of the range specified in the literature. Further, it is relevant to note that the mean of YMB is evaluated as 6.82 with a standard deviation of 1.84, if the values of budget parameters lie between 80 and 100%. The comparable values reported in the literature are between 2 and 9%. Moreover, for L_4 level of maintenance that corresponds to values of budget parameters lying between 80 and 90%, the mean of YMB is evaluated as 10.52% with a standard deviation of 1.02% of ARV. Experience shows that most well-maintained enterprises having modern maintenance management outlook do resort to maintenance budgets of this order. Briefly, the results of Table 2

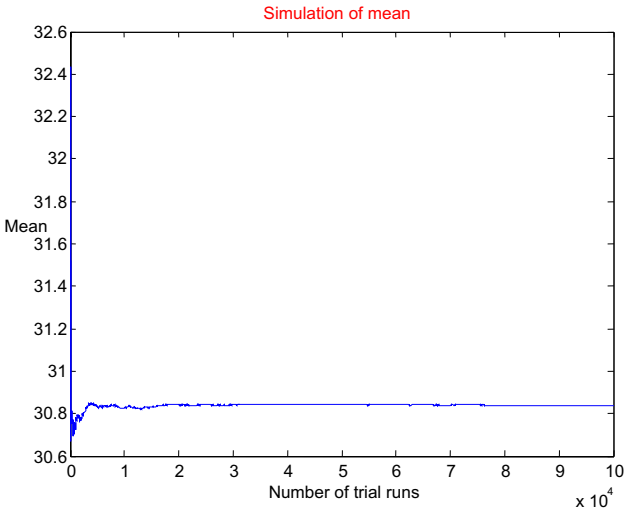


Fig. 8 Mean of YMB Versus number of simulations for L_6 level of maintenance

indicate that management must resort to corrective actions, if the values of budget influencing parameters drop below 80%.

The next section discusses the distribution followed by YMB function and the confidence interval (CI) estimation of mean and variance of YMB using sample data generated by MCS through inferential statistics.

6 Distribution of YMB function

Inferential statistics was applied to find the appropriate distribution and evaluate the interval estimate of maintenance budget parameters.

6.1 Budget function distribution

In this sub-section, YMB distribution is identified by generating a random data set of YMB using MCS.

6.2 Identification of distribution

The best distribution was identified by fitting theoretical to the generated data set. This was based on statistically significant sample size of 100 with margin of error in the range of 9–10%, which is a norm in industry. For illustration purposes, reference may be made to Table 3.

Exponential, Weibull, lognormal and normal distributions were tested for the best fit with the generated data set. Index of fit values of '*r*', were found to be 0.990, 0.974, 0.989 and 0.936 for normal, Weibull, lognormal and exponential distributions, respectively, implying that Normal distribution is the best fit. Illustrative plots for this distribution are only shown in Figs. 9, 10, 11 and 12.

Similarly, the best-fit distributions for other data sets were evaluated for other ranges of budget influencing variable and the results are shown in Table 4.

Table 3 YMB data for maintenance level 'L_2'

S. No.	YMB	S. No.	YMB	S. No.	YMB	S. No.	YMB	S. No.	YMB
1.	8.02	21.	7.55	41.	5.09	61.	4.86	81.	6.01
2.	4.89	22.	7.39	42.	7.73	62.	9.46	82.	8.36
3.	5.082	23.	3.83	43.	7.59	63.	4.38	83.	4.99
4.	5.282	24.	5.18	44.	5.40	64.	7.41	84.	8.88
5.	6.42	25.	7.76	45.	8.53	65.	8.02	85.	7.55
6.	7.05	26.	7.46	46.	5.28	66.	4.89	86.	7.39
7.	8.63	27.	5.49	47.	5.65	67.	5.08	87.	3.83
8.	6.47	28.	9.37	48.	8.67	68.	5.28	88.	5.18
9.	9.35	29.	4.89	49.	6.10	69.	6.44	89.	7.76
10.	5.95	30.	7.49	50.	9.33	70.	7.05	90.	7.46
11.	7.41	31.	10.04	51.	6.76	71.	8.63	91.	5.46
12.	4.94	32.	4.62	52.	6.27	72.	6.47	92.	9.37
13.	9.02	33.	6.25	53.	5.56	73.	9.35	93.	4.89
14.	11.4	34.	7.20	54.	7.26	74.	5.95	94.	7.49
15.	7.68	35.	7.24	55.	8.09	75.	7.41	95.	10.04
16.	5.9	36.	6.34	56.	3.92	76.	4.94	96.	4.62
17.	6.01	37.	9.51	57.	6.02	77.	9.02	97.	6.25
18.	8.36	38.	6.59	58.	3.80	78.	11.4	98.	7.20
19.	4.50	39.	7.13	59.	7.16	79.	7.68	99.	7.24
20.	8.88	40.	3.92	60.	9.51	80.	5.93	100.	6.34

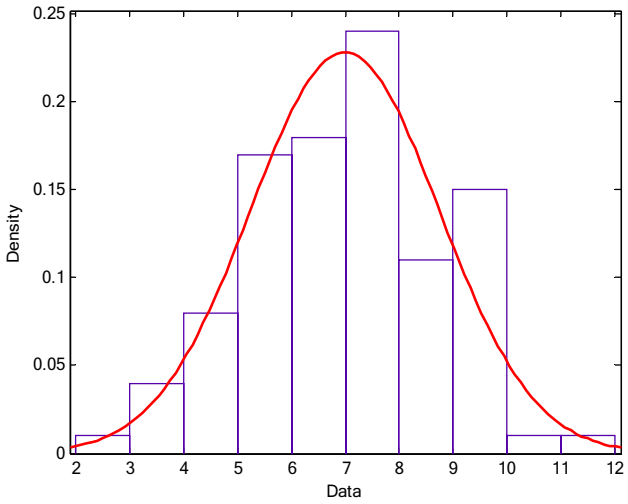


Fig. 9 Probability density function (Normal) for L_2 level of maintenance

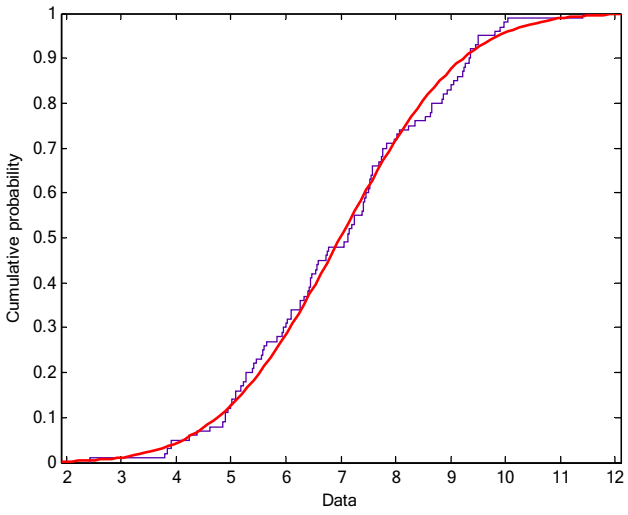


Fig. 10 Cumulative density function (Normal) for L_2 level of maintenance

Goodness-of-fit test was carried out for validating the best fit, which is described in the next sub-section.

6.3 Goodness-of-fit test

This sub-section deals with testing of best-fit distribution identified in Sect. 6.2 using inferential statistics, i.e., Goodness-of-fit test. (Ross 1987) described the Chi-square goodness-of-fit test, which is a process of identification of theoretical distribution. The test is carried on data generated for values of maintenance budget parameters in the range of 80–100% and the test procedure is as follows.

The hypothesis is designed as:

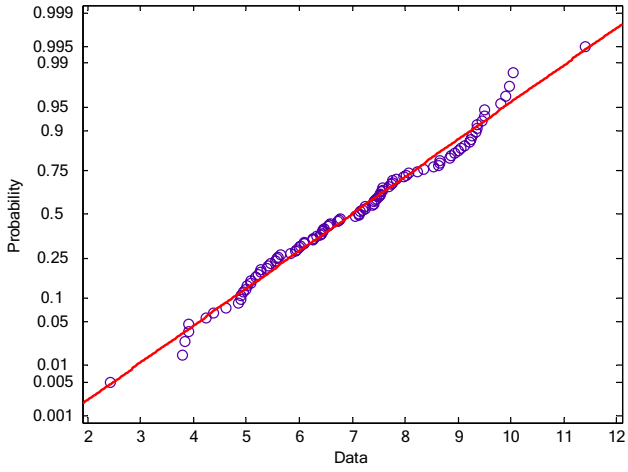


Fig. 11 Probability plot (Normal) for L_2 level of maintenance

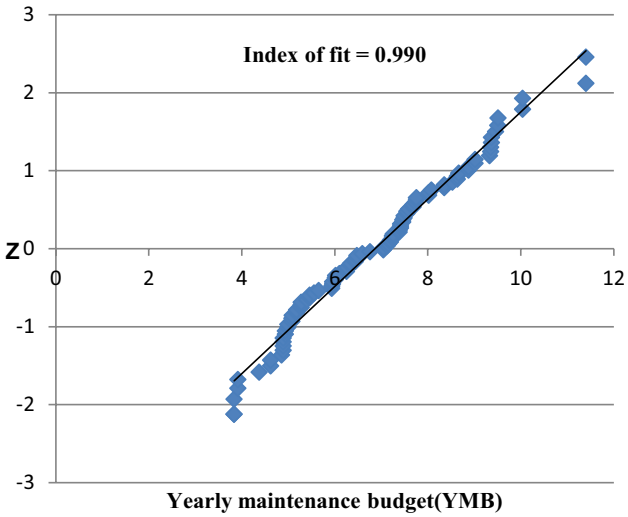


Fig. 12 Least-square curve-fitting plot (Normal) for L_2 level of maintenance

H_0: The maintenance budget data follow the normal distribution

H_1: The maintenance budget data do not follow normal distribution

To carry out the Chi-square test, a significance level of 0.05 is assumed. The test statistic value is calculated using the sample data of Table 3 and its value is 6.47. The critical value of Chi square from the statistical table is noted as 11.07 (Levin et al. 2009). It is concluded that mean of YMB follows normal distribution. On similar lines, sample data of YMB for other ranges of YMB parameters were also tested for best-fit distribution and the results are summarized in Table 4.

Further, confidence interval estimates for mean and variance of YMB were evaluated and are shown in Table 5 for

different maintenance levels defined earlier in Table 2. Also, Table 6 shows the results of hypothesis testing of mean and variance based on significance level of 0.05.

7 Practical implications

The results discussed in Sects. 5 and 6 identify the requirement of budget for different maintenance level. For world-class practices and L_1 to L_3 level of maintenance, the budget value falls between 2 and 9%. This indicates that the maintenance department is following optimum processes of maintenance. If the maintenance budget ranges from 10 to 20% as in the case of L_4 and L_5 level, it shows that unnecessary maintenance expenditure reduces the profitability of the organization. To face the above challenges, organizations must constantly pursue maintenance excellence through highly structured maintenance programs. Over 20% of the budget value indicates that yearly maintenance expenditure is so high at L_6 maintenance level and most of the systems in plant should be replaced by a completely new one at regular intervals. Maintenance engineer shall decide the value of maintenance parameters based on plant-specific conditions and may use the suggested methodology to estimate the yearly maintenance budget. Corrective actions must be taken by maintenance engineers if budget influencing parameter

Table 4 Distribution and goodness-of-fit test

	Range of maintenance parameters for various maintenance levels	Best-fit distribution (based on plots and index of fit values)	Result of goodness-of-fit test (Y for null hypothesis) (N for alternate hypothesis)
World-class practices	Range of parameters as per Table 1	Normal	Y
Maintenance levels	L_1	Normal	Y
	L_2	Normal	Y
	L_3	Normal	Y
	L_4	Normal	Y
	L_5	Normal	Y
	L_6	Not fitting into any of the four chosen distributions	N

Table 5 Confidence interval estimates for mean and variance

	Range of maintenance parameters for various maintenance levels	95% CI estimate for Mean	95% CI estimate for Variance
World-class practices	Range of parameters as per Table 1	(1.62, 1.82)	(0.22, 0.38)
Maintenance levels	L_1	(2.99, 3.31)	(0.42, 0.73)
	L_2	(6.51, 7.23)	(2.31, 4.04)
	L_3	(6.36, 7.64)	(6.44,11.27)
Maintenance levels	L_4	(10.45,10.85)	(0.93, 1.63)
	L_5	(19.28,19.78)	(1.27,2.23)
	L_6	(30.67, 31.33)	(1.91,3.35)

Table 6 Hypothesis test for population parameters

	Range of maintenance parameters for various maintenance levels	Test for mean at 5% significance level	Test for variance at 5% significance level
World-class practices	Range of parameters as per Table 1	H_0 accepted	H_0 accepted
Maintenance levels	L_1	H_0 accepted	H_0 accepted
	L_2	H_0 accepted	H_0 accepted
	L_3	H_0 accepted	H_0 accepted
	L_4	H_0 accepted	H_0 accepted
	L_5	H_0 accepted	H_0 accepted
	L_6	H_0 rejected	H_0 rejected

values fall below 80%. The methodology recommended in this work will be a valuable guideline to maintenance engineer in deciding YMB required for plant maintenance. It also provides indications for maintenance system rectifications and avoids expert recommendations on maintenance budget allocation.

8 Conclusions and future scope

Estimation of YMB of a plant system is an important managerial decision, which to date has been primarily based on historical data, expert opinions, assuming it as some percentage of the capital cost, etc. all of which are non-scientific. Literature revealed that there are no scientific guidelines for such a critical operational decision. In this work, Monte Carlo simulations were used to evaluate YMB.

The work has attempted to present the ranges of YMB as a % of asset replacement value for different maintenance levels practiced in the plant systems. The study concludes that for 95% CI estimate for mean of YMB for world-class maintenance practices lies between 1.62% and 1.82% of ARV. For Level 1 maintenance standard that corresponds to values of maintenance parameters upward of 90%, it lies between 2.99% and 3.31% of ARV. For Level 2 maintenance standards that correspond to values of maintenance parameters upwards of 80%, it lies between 6.51% and 7.23% of ARV. Simulation results also show that operational managers must intervene for systemic corrections, if the budget influencing parameters fall below 80% as these will enhance the YMB significantly and might severely hamper the operational costs in plant systems. The values evaluated in this paper may be realistically validated on the ground by practicing managers.

Understanding the complex nature of identification and quantification of maintenance parameters to carry out the budget estimation is a challenging issue to be resolve effectively. Artificial intelligence (AI) may assist in achieving this goal. In future work, neural network, evolutionary approaches or hybrid method of AI techniques may be implemented on this study to obtain more accurate budget estimate of a plant.

References

Alsyouf I (2009) Maintenance practices in Swedish industries: survey result. Int J Prod Econ 121(1):212–223

Bahr C, Lennerts K (2010) Quantitative validation of budgeting methods and suggestion of a newcalculation method for the determination of maintenance costs. J Facil Manag 8(1):47–63

Borgonovo E, Marseguerra M, Zio E (2000) A Monte Carlo methodological approach to plant availability modeling with maintenance, aging and obsolescence. Reliab Eng Syst Saf 67(1):61–73

Cadini F, Gioletta A (2016) A Bayesian Monte Carlo based algorithm for the estimation of small failure probabilities of system affected uncertainties. Reliab Eng Syst Saf 153:15–27

Campbell D (1995) Uptime: strategies for excellence in maintenance management. Productivity Press, Portland

Cantoni M, Marseguerra M, Zio E (2000) Genetic algorithm and Monte Carlo simulation for optimal plant design. Reliab Eng Syst Saf 68(1):29–38

Chan FTS, Lau HCW, Ip RWL, Chan HK, Kong S (2005) Implementation of total productive maintenance: a case study. Int J Prod Econ 95(1):71–94

Crespo MA, Iung B (2007) A structured approach for the assessment of system availabilityand reliability using Monte Carlo simulation. J Qual Maint Eng 13(2):125–136

Dekker R (1996) Applications of maintenance optimization models: a review and analysis. Reliab Eng Syst Saf 51(3):229–240

Dordevic M, Petrovic S, Peric S, Spasic D (2012) Maintenance cost budgeting optimization. In: Proceedings 5th international scientific conference on defensive technologies (OTEH, Belgrade, Serbia

Ferreira RS, Barroso LA, Feinstein CD, Borges CLT (2013) A non-linear programming approach to maintenance budgeting for multi-component systems. IEEE Power & Energy Society General Meeting, Vancouver, BC, pp 1–5

Gulati R, Smith R (2009) Maintenance and reliability best practices. Industrial Press Inc, New York

Gupta P (2017) Maintenance budget estimation for a Particle Accelerator System under its contextual conditions: a case study. Int J Syst Assur Eng Manag 88(Suppl. 2):S1143–S1153

Gupta S, Gupta P (2017) Evaluating maintenance budget for an AC plant under its contextual conditions. J Facil Manag 15(1):76–89

Gupta P, Gupta S, Gandhi OP (2014) Annual maintenance budget estimation for a plant system—Digraph model and matrix approach. J Qual Maint Eng 20(2):193–210

Hall P, Strutt E (2003) Probabilistic physics of failure models for component using Monte Carlo simulation and Weibull analysis: a parametric study. Reliab Eng Syst Saf 80(3):233–242

Janssen H (2013) Monte Carlo based uncertainty analysis: sampling efficiency and sampling convergence. Reliab Eng Syst Saf 109:123–132

Jurkat WB, Ryser HJ (1966) Matrix factorization of determinants and permanents. J Algebra 3:1–27

Kayrbekova D, Barbadi A, Markeset T (2011) Maintenance cost evaluation of a system to be used in Arctic Conditions: a case study. J Qual Maint Eng 17(4):320–336

Komonen K (2002) A cost model of industrial maintenance for profitability analysis and benchmarking. Int J Prod Econ 79(1):15–31

Kumar VNA, Gandhi OP (2011) Quantification of human error in maintenance using graph theory and matrix approach. Qual Reliab Eng Int 27(8):1145–1172

Kumar G, Varghese JP (2018) Optimum preventive maintenance policy for a mechanical system using Semi-Markov approach and Golden section search technique. In: IEEE International conference on industrial engineering and engineering management, Bangkok, Thailand, Dec 16–19, 2018, pp 232–236

Kumar G, Jain V, Gandhi OP (2014) Steady-state availability analysis of repairable mechanical systems with opportunistic maintenance by using Semi-Markov Process. Int J Syst Assur Manag 5(4):664–678

Kumar G, Jain V, Gandhi OP (2018) Availability analysis of mechanical systems with condition based maintenance using semi-Markov and evaluation of optimal condition monitoring interval. J Ind Eng Int 14(1):119–131

Levin DM, Stephan DF, Krehbiel TC, Berenson ML (2009) Statistics for managers. PHI Learning Private Limited, New Delhi

Loganathan MK, Gandhi OP (2016) Cost minimization of manufacturing systems using PSO under reliability constraint. Int J Syst Assur Eng Manag 7(1):47–61

Loganathan MK, Neog SS, Rai S (2018) Process safety and performance improvement in oil refineries through active redundancy and risk assessment method: a case study. In: Proceedings of 2018 IEEE international conference on industrial engineering and engineering management, pp 98–102

Loganathan MK, Bezbaurah I, Gandhi OP, Borah RC (2018b) Criticality analysis of wind turbine energy system using fuzzy digraph models and matrix method. In: Haugen S et al (eds) Safety and reliability—safe societies in a changing world. Taylor & Francis Group, London, pp 1727–1733. ISBN 978-0-8153-8682-7

Louit D, Pascual R, Banjevic D, Jardine A (2011) Optimization model for critical parts inventories—a reliability approach. J Oper Res Soc 62(1):992–1004

Marquez A, Heguedas A, Iung B (2004) Monte Carlo—based assessment of system availability A case study for cogeneration plants. Reliab Eng Syst Saf 88(3):273–289

Marseguerra M, Zio E (2000) Optimizing maintenance and repair policies via combination of genetic algorithms and Monte Carlo simulation. Reliab Eng Syst Saf 68(1):61–73

Marseguerra M, Zio E, Podofillini L (2002) Condition-based maintenance optimization by means of Genetic algorithm and Monte Carlo simulation. Reliab Eng Syst Saf 77(2):151–165

Muchiri P, Pintelon L, Gelders L, Martin H (2011) Development of maintenance functionperformance measurement framework and indicators. Int J Prod Econ 131(1):295–302

Nahas N, Khatab A, Ait-Kadi D, Nourelfath M (2008) Extended great deluge algorithm for the imperfect preventive maintenance optimization of multi-state systems. Reliab Eng Syst Saf 93(11):1658–1672

O'Leary N, Ahrend A, McDonaeh A, O'Brien B, Shallo L (2018) A partial budgeting tool to assess investments in Precision Dailry Technologies. Sustainable meat and milk production from grasslands. In: Proceedings of 27th general meeting of the European Grassland Federation, Cork, Ireland, 17–21 June 2018, pp 810–812

Ottoman GR, Nixon WB, Lofgren ST (1999) Budgeting for facility maintenance and repair. I: methods and models. J Manag Eng 15(4):71–83

Pham H, Wang H (1996) Imperfect maintenance. Eur J Oper Res 94(3):425–438

Ramakrishnann M (2016) Unavailability estimation of shutdown system of a fast reactor by Monte Carlo Simulation. J Ann Nucl Energy 90:264–274

Ross SM (1987) Introduction to probability and statistics for engineers and scientists. Wiley, New York

Sarker BR, Faiz TI (2016) Minimizing maintenance cost for offshore wind turbines followingmulti-level opportunistic preventive strategy. Renew Energy 85:104–113

Swanson L (2001) Linking maintenance strategies to performance. Int J Prod Econ 70(3):237–244

Topal E, Ramazan S (2010) A new MIP model for mine equipment scheduling by minimizing maintenance cost. Eur J Oper Res 207(2):1065–1071

Yuzvovich L, Korogodina E, Azisova L (2018) Improving the system of budgeting at Industrial enterprises. SCH Web Conf 50:01015

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Our Heritage

ISSN: 0474-9030

Vol-67-Issue-5-September-2019

Exploring Barriers of Corporate Social Responsibility: Indian SMEs Perspective

Dr. Mohd Imran Khan[a], Shahbaz Khan[b*], Mohd Shuaib[c]

[a]Mittal School of Business, Lovely Professional University, Jalandhar-Delhi, G.T. Road, Phagwara, Punjab (INDIA)-144411.

[b]Department of Mechanical Engineering, Faculty of Engineering and Technology, Jamia Millia Islamia, New Delhi-110025, India

E-Mail: shahbaz.me12@gmail.com

[c]Mechanical Engineering, Delhi Technological University Delhi-110042

*Corresponding Author

Abstract

Corporate social responsibility (CSR) aims at improving the well-being of the firm's stakeholders through voluntary business practices in order to generate revenue. Businesses have comprehended to thrive they need to co-relate their business model with social performance. This result in the development of concept corporate social responsibility (CSR) which focuses on incorporating transparency and accountability towards society. Indian Small and Medium Enterprises (SMEs) are the biggest employment generator, however they are seemed to least concern about the CSR issues. The objective of this research paper to highlight the barrier faced by the Indian SMEs towards implementation of the CSR practices in their business operations. This paper concludes with the statement that CSR activity can be used as strategic tool to create a differential advantage over competitors.

Keywords: Corporate Social Responsibility, Indian manufacturing small and medium enterprises (SMEs), Barriers

1.0 Introduction

Braune et al., (2019) argued business organizations needs to follow the ethical norms as well as societal norms in addition to revenue generation and adhering to the laws and regulation. Term Corporate Social Responsibility (CSR) aims at increasing the transparency as well as openness

Copyright ⓒ 2019Authors

ISSN: 0474-9030

Vol-67-Issue-5-September-2019

of the business operations in order to consider the will and expectation of the stakeholders. The European Union defines CSR as "the responsibility of enterprises for their impacts on society". To follow the CSR practices businesses "should have in place a process to integrate social, environmental, ethical human rights and consumer concerns into their business operations and core strategy in close collaboration with their stake holders." The World Business Council for Sustainable Development defines CSR as "continuing commitment by business to contribute to economic development while improving the quality of life of work force and their families as well as of the community and society at large." According to the United Nation Industrial Development Organization, "Corporate social responsibility is a management concept whereby companies integrate social and environmental concerns in their business operations and interactions with their stakeholders." The implementation of CSR is compassionate, or socially motivated, with the objective to serve the society and reduce the negative externalities at the cost of profits. CSR help in achieving the firm's bottom line.

CSR is distinct from the concept of charity, sponsorships of philanthropic activities as this is a strategic business management concept which aims at enhancing the brand value of the company. Above discussion demonstrate that:

- CSR activities is holistic and integrated with core business strategy for addressing social and environmental impacts of businesses.
- CSR needs to address the well-being of all the stake holders not only company's shareholders.
- Philanthropic activities are only a part of CSR, which otherwise constitutes a much larger set of activities entailing strategic business benefits.

Importance of CSR practices have grown to such extent the that they have the capability to erode the brand value of the and legitimacy (Wertherand Chandler 2005).

Due to great penetration of the information technologies, communities are becoming more aware and demanding. Businesses have also realized that license to function in a particular area is not only sanctioned by the government but also by the communities which are affected by the business operations. Robust implementation of the CSR activity will help in reducing the deficit of community development. Several studies (Haddock, and Fraser, 2008; Faisal, 2010) have found CSR activities as a tool to attract, retain as well as motivate talent. As these activities give

Copyright ⓒ 2019Authors

ISSN: 0474-9030

Vol-67-Issue-5-September-2019

a sense of belonging and boost employee morale. CSR activity can enhance the livelihood of the community by incorporating then into the supply chain. Effective CSR activities create a positive image and goodwill and project them as a responsible corporate citizen.

1.1 CSR in India

India is a diverse country in terms of economic development. It has emerged as third largest economy and an important global player, however it is still a home for largest number of people living in absolute poverty with a great proportion of undernourished children. This inequality is due to uneven distribution of the benefits of growth which many believe, is the root cause of social unrest. In general CSR in India is termed as philanthropic activity and done deliberately with limited documentation on activities related to CSR concept (Bhattacharyya & Rahman, 2019).

The idea of CSR is also supported by several religion where it has been intertwined with religious laws. Focussed on "Zakaat", "Sadqa" followed by Muslims is donation from one's earning, which is specifically given to the poor and needy persons. Similarly, Hindus follow the principle of "Dhramada" and Sikhs "Daashant".

CSR in India is still limited in philanthropic space; however, it is moved from building institutions for education as well research and culture for the community (Mukherjee et al., 2018). Moreover, due to globalization and emerging of newer communication technologies stakeholders are demanding which changed the nature of the philanthropic activity from community development to incorporating it as a catalyst for brand development. Now businesses are reporting the CSR practices on their online platforms, annual report as well there are evidence companies are even releasing the CSR reports (Chang et al., 2019).

From the regulation point of view The Companies Act 2013, forced the businesses to perform the CSR activity in its true value by proper documentation as well as releasing annual report. This act has documented the activities which will be counted under CSR and suggested that communities needs to be focal point. It forces the organizations to integrate the CSR activity in core business operation and linked with the performance measure of the organization.

2.0 CSR and SMEs

Functioning of the SMEs is vital for the growth of the Indian economy. The operations of the SMEs are quietly linked with the local community (Murillo & Lozano, 2006). This prompts them

Copyright ⓒ 2019Authors

ISSN: 0474-9030

Vol-67-Issue-5-September-2019

to have a good knowledge of community needs and expectations and helps them to develop the appropriate CSR program which focus on community development (Jenkins, 2006).

2.1 Barriers to CSR Implementation in SMEs

Ten main barriers to the implementation of CSR practices in the business operations of Indian SMEs were identified after reviewing and synthesizing the existing literature on CSR reporting as well as CSR implementation.

2.1.1 Financial Constraints

Several studies have positively linked CSR activities with firm's business performance (ref). however, allocation of budget to tackle the negative impact of the business operations tells different story. Mili et al., (2019) concluded that absence of a robust framework that links financial performance and social outcomes inhibits the promotion of CSR for community development. Moreover, Faisal (2010) also reported that funds for carrying out CSR operations are often inadequate and outcompeted by other projects which guarantee higher return on investments. SME's view allocation of budget for CSR activities as an expense rather than an investment. Major concern of the SME's is to maximize the profit, demotes CSR practices.

2.1.2 Lack of Concern About Reputation

Brand building as well generating goodwill is the major enabler that force the big corporations to implement CSR practices. However, in general, major SMEs, do not posses a brand value of their own moreover, they are branded as a supplier or supply chain partner of large manufacturing companies (Luffarelli et al., 2019). Due to this they are not concerned about the CSR practices as they seem to believe that any negligence on their part is not going to affect them directly.

2.1.3 Lack of Strategic Vision

Lack of top management commitment leads to lack of strategic vision. As per Werther and Chandler, (2005) CSR must be implemented form the highest executive levels in order to gain its full potential. As CSR needs management should think beyond financial performance of the company (Al-Dah, 2019). Khan et al., (2014) suggested that SMEs are implementing CSR only at the operational level and have no strategic vision regarding this. Another reason SMEs are

Copyright ⓒ 2019Authors

ISSN: 0474-9030

Vol-67-Issue-5-September-2019

reluctant to implement CSR as they are more focused towards short-term goals and profits rather than building a credible brand.

2.1.4 Low Willingness to Pay for CSR

In general customers show low willingness to pay for any premium concerning with CSR. Consumers demands that firm should be engaged in CSR practices, however they are not rewarding by changing their purchasing behavior (Campa & Zijlmans, (2019). While few studies suggest that consumers punish the firms who are insincere in their social involvement. SMEs also shows low interest towards paying premium as they are generally involved in B2B industry and they are rarely scrutinized for their involvement in CSR (Castro-González et al., 2019; Huang, & Kumar, 2019).

2.1.5 Lack of Awareness About What Constitutes CSR

Term CSR is an ambiguous term and often used interchangeably with other similar terminologies such as "corporate sustainability", "corporate governance", "business ethics", as well as "corporate governance" (Chiesa & Przychodzen, 2019). Also, unavailability of metric which quantify the benefits for implementing CSR practices in the business operations making the implementation as well as evaluation difficult. CSR indicators often reported by the business firms in their annual report are not standardized.

2.1.6 Reluctance of Partners to Implement CSR

In the agile market business firms are mostly dependent upon the supplier to deliver the desired goods as well as services (McWilliams & Siegel, 2011). However, as per CSR norms, firms need to disclose the supplier's activity related to social as well as environmental issues in addition to their own activity (Yen et al., 2019). Moreover, in global supply chain, supplier may be located in different countries which may have distinct environmental as well as social issues. It is difficult for the parent firm to guide their partners to adhere the particular CSR philosophy.

2.1.7 Lack of CSR Enforcement Mechanism

In general CSR activities are voluntary and it is expected companies will themselves practice and monitor the CSR initiatives. However, manager's incentive to implement and monitor CSR practices are low as their performance is evaluated in terms of financial results (Yuen & Lim, 2016). Prevalent CSR codes does not offer any substantial evidence for transforming business

Copyright ⓒ 2019Authors

ISSN: 0474-9030

Vol-67-Issue-5-September-2019

performance (Williamson, 2006). Studies show companies with proper environmental management system perform same as those without. One of the major problems towards inculcation of CSR practices in business operations is the lack of proper of enforcement mechanism particular at international level (Detienne, and Lewis, 2005). Large-scale corruption in public life coupled with unemployment in India weakens any CSR enforcement mechanism and is a big challenge for SMEs (Agyei-Mensah & Buertey 2019).

3.0 Discussion and Conclusion

SMEs are reluctant to implement CSR practices as they believe that these will not affect their financial performance and hence not allocating budget for the same. Furthermore, absence of evidence which co-relate the relationship between CSR practices and financial performance of the company makes the adaptation more complex. Several previous studies failed to establish the relationship. Lack of understanding particularly at the top level of Indian SMEs regarding the benefit of the CSR aggravate the situation. Proper training to sensitize the top management towards CSR will help in realize its potential benefit .In addition to financial constraints, SMEs also face lack of human resources to implement the CSR, NGOs and government authorities needs to collaborate with the SMEs in order to remove the negative impact of the business operations.

SMEs are failing to address the CSR issues as these codes are in general designed for big corporations. CSR implementation codes needs to be tailored as per the requirement of SMEs and make them more accessible and relevant in order to encourage the SMEs towards increasing the wellbeing of the stakeholders.

REFERENCES

Agyei-Mensah, B., & Buertey, S. (2019). The effect of corruption and culture on corporate social performance: an empirical study. *Social Responsibility Journal*, *15*(8), 1071-1086. doi: 10.1108/srj-12-2017-0271

Al-Dah, B. (2019). Director interlocks and the strategic pacing of CSR activities. *Management Decision*, *57*(10), 2782-2798. doi: 10.1108/md-11-2017-1103

Bhattacharyya, A., & Rahman, M. (2019). Mandatory CSR expenditure and firm performance. *Journal Of Contemporary Accounting & Economics*, *15*(3), 100163. doi: 10.1016/j.jcae.2019.100163

Braune, E., Charosky, P., & Hikkerova, L. (2019). Corporate social responsibility, financial performance and risk in times of economic instability. *Journal Of Management And Governance*, *23*(4), 1007-1021. doi: 10.1007/s10997-019-09476-y

Copyright ⓒ 2019Authors

ISSN: 0474-9030

Vol-67-Issue-5-September-2019

Campa, D., & Zijlmans, E. (2019). Corporate social responsibility recognition and support for the arts: Evidence from European financial institutions. *European Management Journal*, *37*(6), 818-827. doi: 10.1016/j.emj.2019.01.003

Castro-González, S., Bande, B., Fernández-Ferrín, P., & Kimura, T. (2019). Data to model the influence of CSR on consumer behaviors: A process approach. *Data In Brief*, *27*, 104713. doi: 10.1016/j.dib.2019.104713

Chang, W., Amran, A., Iranmanesh, M., & Foroughi, B. (2019). Drivers of sustainability reporting quality: financial institution perspective. *International Journal Of Ethics And Systems*, *35*(4), 632-650. doi: 10.1108/ijoes-01-2019-0006

Chiesa, P., & Przychodzen, W. (2019). Social sustainability in supply chains: a review. *Social Responsibility Journal*, (ahead-of-print). doi: 10.1108/srj-11-2018-0301

Detienne, K. B., and Lewis, L.W. (2005), "The pragmatic and ethical barriers to corporate social responsibility disclosure: The Nike case." *Journal of Business Ethics*, Vol.60, No.4, pp.359-376.

Faisal, M. (2010). Analysing the barriers to corporate social responsibility in supply chains: an interpretive structural modelling approach. *International Journal Of Logistics Research And Applications*, *13*(3), 179-195. doi: 10.1080/13675560903264968

Haddock, J.F., and Fraser, I. (2008), "Assessing corporate environmental reporting motivations: differences between 'close to market' and 'business -to-business' companies", *Corporate Social Responsibility and Environmental Management*, Vol.15, No.3, pp.140-155

Huang, Y., Do, M., & Kumar, V. (2019). Consumers' perception on corporate social responsibility: Evidence from Vietnam. *Corporate Social Responsibility And Environmental Management*. doi: 10.1002/csr.1746

Jenkins, H. (2006), "Small business champions for corporate social responsibility", *Journal of Business Ethics*, *67(*3), 241-256.

Luffarelli, J., Markou, P., Stamatogiannakis, A., & Gonçalves, D. (2019). The effect of corporate social performance on the financial performance of business- to- business and business-to- consumer firms. *Corporate Social Responsibility And Environmental Management*. doi: 10.1002/csr.1750

McWilliams, A., and Siegel, D.S. (2011), "Creating and capturing value strategic corporate social responsibility, resource-based theory, and sustainable competitive advantage", *Journal of Management*, *37* (5),1480-1495

Mili, M., Gharbi, S., & Teulon, F. (2019). Business ethics, company value and ownership structure. *Journal Of Management And Governance*, *23*(4), 973-987. doi: 10.1007/s10997-019-09475-z

Mukherjee, A., Bird, R., & Duppati, G. (2018). Mandatory Corporate Social Responsibility: The Indian experience. *Journal Of Contemporary Accounting & Economics*, *14*(3), 254-265. doi: 10.1016/j.jcae.2018.06.002

Murillo, D., and Lozano, J.M. (2006), "SMEs and CSR: An approach to CSR in their own words", *Journal of Business Ethics*, 67(3), 227-240.

Copyright ⓒ 2019Authors

ISSN: 0474-9030

Vol-67-Issue-5-September-2019

Werther JR, W.B., and Chandler, D. (2005), "Strategic corporate social responsibility as global brand insurance", *Business Horizons*, *48*(3), 317-324.

Williamson, D., Lynch-Wood, G., and Ramsay, J. (2006), "Drivers of environmental behaviour in manufacturing SMEs and the implications for CSR", *Journal of Business Ethics*, *67*(3), pp.317-330.

Yen, M., Shiu, Y., & Wang, C. (2019). Socially responsible investment returns and news: Evidence from Asia. *Corporate Social Responsibility And Environmental Management*. doi: 10.1002/csr.1833

Yuen, K., & Lim, J. (2016). Barriers to the Implementation of Strategic Corporate Social Responsibility in Shipping. *The Asian Journal Of Shipping And Logistics*, *32*(1), 49-57. doi: 10.1016/j.ajsl.2016.03.006

Khan, M., Khan, U., & Haleem, A. (2014). Corporate Social Responsibility: Modelling of Critical Factors using an Integrated ISM and Fuzzy-Micmac approach. *International Journal Of Global Business And Competitiveness*, *9*(1), 1-13.

Copyright ⓒ 2019Authors

The current issue and full text archive of this journal is available on Emerald Insight at:
www.emeraldinsight.com/2050-7003.htm

Gratitude and leadership in higher education institutions

Exploring the mediating role of social intelligence among teachers

Naval Garg and Shilpika Gera
Delhi Technological University, Delhi, India

Received 20 September 2019
Revised 12 November 2019
Accepted 12 November 2019

Abstract

Purpose – The purpose of this paper is to examine the relationship between gratitude and transformational leadership among university teachers. The study also intended to investigate the mediating effect of social intelligence (SI) on this relationship between gratitude and leadership.
Design/methodology/approach – Correlation and regression are used to analyze the relationship between gratitude, SI and leadership. Mediating effect of SI is accessed using Baron and Kenny's (1986) method and bootstrapping technique.
Findings – Findings reveal positive relationship between gratitude and teacher leadership. SI is also found to mediate partially. The paper also discusses practical implications, limitations and scope for future researchers.
Originality/value – This is one of the pioneer studies that explore the role of gratitude in leadership development. Findings of the study will guide educational administrators to become more effective and efficient leaders.

Keywords Mediation, Gratitude, Educational leadership, Social intelligence
Paper type Research paper

Introduction

Teachers and leaders have a profound impact on every aspect of personal and professional life. And a teacher leader combines both influential words together. Thus, a teacher leader is bound to have more significant influence on students, and academic settings. Acknowledging the importance of developing leaders among teachers, National Education Association of USA initiated "The Teacher Leadership Initiative" to develop and promote leaders within educationists. In India, "School Leadership Development Program" of the Ministry of Human Resource Management and "Faculty Induction Program" of University Grant Commission are aimed at developing future leaders for the Indian education system. Available literature denotes educational leaders with several nomenclature such as program coordinator, dean, head of the department, lead teacher, department chair and mentor (Neumerski, 2013; Mangin and Stoelinga, 2008). Traditionally, teachers were more focused in teaching and research work, but emergence of the concept of teacher leader has diversified their domain of performance.

Different research scholars, including Silva *et al.* (2000), Crowther *et al.* (2009) and Spillane (2006) discussed several scopes for a teacher leader. Silva *et al.* (2000) described three major areas of engagement for teacher leaders. First, teachers are supposed to take managerial responsibilities like dean, head of departments and vice chancellors. A teacher is well aware with the academic matters and thus, he/she can empathize and take appropriate decision. Second, a teacher leader is required to take responsibilities of growth and development of academic through curriculum development, course designing, streamlining student learning and encouraging academia-industry collaboration. And at last, others look at teacher leader as role model. They are supposed to encourage their colleagues and other teachers to adapt and practice the experimented strategies of better and effective learning. These new roles allow teacher leaders to bring school-wide, society-wide and nation-wide positive changes.

Journal of Applied Research in Higher Education
© Emerald Publishing Limited
2050-7003
DOI 10.1108/JARHE-09-2019-0241

Teacher leadership was reported to have profound impact on several academic indicators including student's performance, school improvement (Crowther *et al.*, 2009), higher academic and social success, high morale of faculty, staff and other stakeholders enhanced academic support and administrations creativity and work involvement greater support from parents, peers and other teachers (Froh *et al.*, 2011) and increased sense of connection and owner of school (McCullough *et al.,* 2001). Although benefits of teacher leadership were adequately explored, the factors leading to development of teacher leaders are still scare. This study proposes gratitude as one of the factors leading to development of teacher leadership.

The proposed association between gratitude and leadership is theoretically premised on Fredrickson's (2004) broaden-and-build theory of positive emotions. It is being proposed that gratitude, like other positive emotions of love, joy, pride and contentment, may broaden thought and action repertoire through improving their social intelligence (SI). This broadened SI may build leadership resources among University teachers. Accordingly, the present study proposes that SI will mediate the relationship between gratitude and leadership. The present research paper is broadly divided into four parts. First, the concept of gratitude, SI and leadership is discussed in brief. Second, theoretical relationships between gratitude, SI and leadership are explained. Furthermore, the relationship between gratitude and leadership is investigated. Also, mediating effect of SI is explored. And in the last part of the paper, theoretical and practical implications of the study, limitations of the study and scope for future researchers are discussed.

Theoretical framework and hypotheses

Gratitude and its constructs

Gratitude has been conceptualized in many ways such as a moral virtue, an attitude, an emotion, a personality trait and a coping response (Emmons and McCullough, 2003). Defining it at the trait level, Wood *et al.* (2010) stated that gratitude is a life orientation toward noticing and appreciating the positivity in life. This definition highlighted gratitude as a personality trait that leads to observing and appreciating the good things in life. It has also been defined as an emotion which occurs on receiving a perceived valuable, costly and selfless help from others (Wood *et al.*, 2008). These helps and benefits may be material or non-material. Gratitude involves two things: recognition of the good to oneself by the individual and this good must be attributed to an external source. McCullough *et al.* (2001) posited that gratitude has a "moral affect" such that when people (beneficiaries) feel grateful for receiving a benefit, they respond and act in ways that promote the well-being of their benefactors as well as of other people in general. This highlights gratitude's role in "upstream reciprocity." McCullough *et al.* (2001) identified three moral functions of gratitude: gratitude as moral barometer, gratitude as moral motive and gratitude as moral reinforce. Defining it as a positive emotion that creates an "upward spiral" toward enhanced fulfillment and well-being of individuals, Fredrickson (2004) explains that gratitude widens people's ways of thinking and engaging in creative actions that might be benefitting for not just their benefactors but for other people as well. This "pay it forward" attitude helps grateful people in building many personal resources such as friendships, social bonds and interpersonal relationships as well as developing intellectual resources such as knowledge, new skills, new information and creativity.

Leadership

Burns (1978) gave two categories of leadership – transactional leadership and transforming leadership. Bass (1990) later replaced the word transforming by "transformational leadership." Transactional leadership is exchange-based relation based on existing values and motivations of leaders and followers and Bass (1990) identified transactional leadership behaviors such as contingent reward and management-by-exception, however,

transformational leadership seeks to change the existing values and motivations and includes behaviors such as idealized influence, inspirational motivation, intellectual stimulation and individualized consideration as identified by Bass (1990). Carrying forward Burns (1978) and Bass' (1990) thoughts on shifting the focus from traditional transactional models based on leader–follower exchange relationship to more new age transformational and inspirational leadership models, Avolio *et al.* (2009) emphasized the need for this "new-genre leadership" style characterized by a charismatic, visionary, inspirational and symbolic leader with emphasis on followers' intellectual stimulation and individualized attention. Such leaders facilitate followers' motivations and aspirations, such that the followers identify with the leader's vision and are motivated to high levels of performance (Avolio *et al.*, 2009; Bass, 1990). This is particularly important at the higher education level where students look up to their teachers and are strongly guided by this mentorship. Vinger and Cilliers' study identified that transformational leaders were people oriented, caring and empathizing toward their followers, motivated them and had a passion for change. They posited that such leaders had the ability to manage cultural diversity with regard to their followers and engaged in adaptive communication and open-door policy. Hao and Yazdanifard have linked effective leadership with positive change and innovative ideas and also with followers' trust, positive organizational culture and a clear vision that facilitates learning, teamwork and communication. The authors posit that effective leadership involves providing clear direction to the followers, increasing their commitment to their work and also having a clear vision that the followers identify with.

Social intelligence

Formally introduced by John Dewey in 1909, the term SI has been defined in different ways by researchers. Goleman and Boyatzis (2008) defined SI as a relationship-based construct and as a set of interpersonal competencies built on specific neural circuits that inspire others to be effective. Equating SI to social competence, Marlowe (1986) defined SI as the ability to understand the feelings, thoughts and behaviors of persons, including one, in interpersonal situations and to act appropriately upon that understanding. Eshghi *et al.* (2013) observed commonness of two components in all definitions of SI: the awareness of others and their response and adaptation to others and the social situations. SI encompasses establishing relationships with others, ability to judge about others' feelings and temperaments, ability to sympathize and empathize with others and being able to decode non-verbal signs (Garg, 2017c). In the book titled *Social Intelligence*, Goleman and Boyatzis (2008) have characterized socially intelligent people as having: awareness and sensitivity toward other people and social facility, i.e. what we do with that awareness. Silvera *et al.* (2001a, b), while constructing the Tromso SI Scale (TSIS), explained SI as a multifaceted construct comprising of three factors: social information processing, social awareness and social skills; and pointed out that components based on reasoning skills and knowledge were more closely related to general intelligence while components based on social setting performance were more closely related to personality traits like extraversion.

Gratitude, leadership and social intelligence

Researchers have studied the effects of many positive emotions on leadership in the past. Howells (2013) established in his study that gratitude had a positive impact on school teachers and the overall classroom environment, elaborating that gratitude among teachers enhanced teacher–student relationship. Chan studied the impact of gratitude on subjective well-being of teachers. Algoe *et al.* (2008) observed that grateful leaders further express gratitude toward followers therefore communicating a pro-social, other-regarding outlook. Ritzenhöfer *et al.* (2017) found that grateful leaders created positive situations that resulted in followers rating such leaders high in trustworthiness, benevolence and integrity (Garg, 2018).

Waters and Stokes (2015) studied the impact of two elements, emotion gratitude and action gratitude on school leaders and found a positive impact on their "within well-being" as well as "relational well-being." Ritzenhöfer *et al.* (2017) studied the influence of two positive emotions, namely, pride and gratitude, measured as dispositional tendencies, on organizational leaders with respect to the pro-social behaviors exhibited by them (social justice and altruism). Apart from concluding that grateful leaders recognize the contributions of others while attributing success, the findings of the study also indicated that leaders possessing grateful emotion toward organizational members could regulate their self-pride and hence could ably focus on the well-being of others (Garg, 2017a; Garg *et al.*, 2019). The author posited that a leader's ability to balance self-interests with the organizational interests could be enhanced by a grateful outlook. Hence, we hypothesize that gratitude positively enhances leadership skills:

H1. Gratitude positively influences leadership skills of teachers.

Although a vast research has been done positively linking gratitude with emotional intelligence, not much has been researched about its possible linkage with SI. Since SI involves an ability to develop sensitivity and awareness toward other people (Goleman and Boyatzis, 2008; Garg, 2017b), and since, gratitude through its broadening mechanism expands people's cognitions such that they become more open to processes like exploration and integration and hence build social skills, new bonds and support as well as develop new knowledge and skills (Fredrickson, 2004), we assume that gratitude positively enhances SI among individuals:

H2. Gratitude positively enhances SI.

Goleman and Boyatzis (2008) emphasized on SI skills for effective leadership. They even went on to highlight that gender differences impacted SI skills needed for outstanding leadership as women tended to rank better at sensing other people's emotions than men. They developed the emotional and social competency inventory involving 360-degree evaluation in which bosses, peers, direct reports, clients and at times even family members assess a leader according to seven SI qualities – empathy (understanding and sensitiveness), attunement (to others' moods and listening ability), organizational awareness (appreciation of organizational culture and awareness of social setting), influence (persuading others and getting support), developing others (coaching and providing feedback), inspiration (instilling vision and leading) and teamwork (encouraging cooperation). The authors also found performance gaps between socially intelligent and socially unintelligent leaders. Hence, we hypothesize that SI positively influences leadership skills:

H3. SI positively impacts leadership skills.

Based on Fredrickson's (2004) work on gratitude's ability to broaden people's mindsets and various thought-action tendencies which help them build various personal, psychological and intellectual resources for themselves, we propose that gratefulness broadens people's cognitions such that it uplifts their emotional and SI skills, which are a prerequisite to building constructive and successful leaders. Hence, we hypothesize that gratitude builds effective leadership through the presence of SI:

H4. Social intelligence mediates the relationship between gratitude and leadership.

Based on above discussed premises, following mediation model is proposed (please refer to Figure 1).

Methodological framework

Objectives, sample and procedures

The primary objective of this study was to examine the relationship between gratitude and transformational leadership among university teachers. The study also intended to investigate

the mediating effect of SI on this relationship between gratitude and leadership. The sample comprised of 150 teachers working in various higher institutions across Delhi-NCR. The data were collected using a structured questionnaire with random sampling. Respondents were approached either personally or through e-mails. Only those teachers who were willing to participate were included and they were assured of confidentiality of information shared with us. They were also assured for academic use of collected data. The composition of respondents was: 49 percent male (73) and 51 percent female (77). Also, 81 respondents belonged to "below 30 years" age category and other 39 and 18 belonged to 31–40 and 41–50 age group, respectively. In addition, eight faculty members were above the age of 50 years.

Measures

Gratitude. Gratitude was measured using the unidimensional Gratitude Questionnaire-6 (GQ-6), developed by McCullough *et al.* (2001). The scale consists of six items measured on a seven-point scale ranging from 1(strongly disagree) to 7(strongly agree), with two items being reverse scored. Examples of GQ-6 include "If I had to list everything that I felt grateful for, it would be a very long list" and "Long amounts of time can go by before I feel grateful to something or someone" (reverse scored).

Leadership. Leadership was accessed using Multifactor Leadership Questionnaire form 6S, developed by Bass and Avolio (1992). The questionnaire consisting of 21 items measures the leadership style on a five-point scale ranging from 0 (not at all) to 4 (frequently, if not always). Examples of items included in the questionnaire were "I make others feel good to be around me," "Others have complete faith in me" and "Others are proud to be associated with me."

Social intelligence. SI was examined using the self-measure TSIS, developed by Silvera *et al.* (2001a, b) and consisting of 21 items, divided into three factors, namely, social information processing (SI Factor 1), social skills (SI Factor 2) and social awareness (SI Factor3); each factor consisting of seven items. Each item is measured on a five-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). Social information processing measures the leader's ability to understand human relations through verbal and non-verbal messages and to empathize with others. Examples include "I know how my actions will make others feel" and "I understand others' wishes." Social skills measure basic communication skills including active listening, establishing, maintaining and breaking up a relationship. Examples include "I fit in easily in social situations" and "I have a hard time getting along with other people" (reverse scored). Social awareness measures a leader's awareness and sensitivity to behave according to the situation, time and place. All items in this factor are reverse scored. Examples include "Other people become angry with me without me being able to explain why" and "I have often hurt others without realizing it."

Reliability, validity and common method variance (CMV)

Prior to statistical analysis, it is first important to check the reliability and validity of the data. Reliability refers to consistency of results and is defined as the degree to which consistent results are produced in spite of repeated measurements. For confirmation of reliability, the

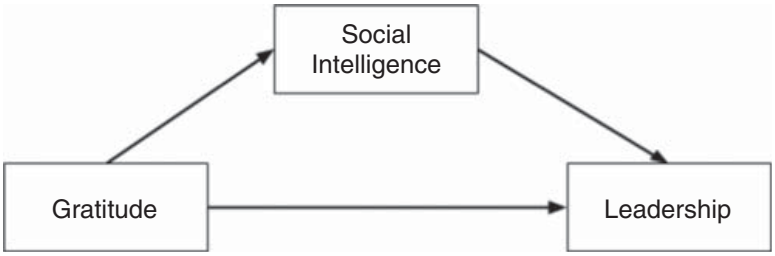


Figure 1.
Proposed mediation model

Cronbach's α value (internal consistency) should be greater than 0.7 (George and Mallery, 2003). Cronbach's α is the mean value of all possible split-half correlations for a set of items. Table I shows acceptable values of all αs (gratitude = 0.798, SI = 0.839, leadership = 0.908). It is further suggested to calculate composite reliability (CR) along with Cronbach's α (Cronbach and Shavelson, 2004). The value of CR should also be greater than 0.70 (reported in Table I). Furthermore, validity is defined as a degree to which instruments measure the variable they were intended to access. Convergent validity was investigated with the help of average variance extracted (AVE), whose values should be greater than 0.50 (Fornell and Larcker, 1981). Following mathematical formulas were used to calculate CR and AVE. Table I represents acceptable values of both CR and AVE, therefore it confirms both reliability and validity:

$$\mathrm{CR} = \left(\sum \lambda_i\right)^2 \Big/ \left(\sum \lambda_i\right)^2 + \left(\sum 1-\lambda_i^2\right),$$

where λ is standardized factor loading for item:

$$\mathrm{AVE} = \left(\sum \lambda_i^2\right) \Big/ N,$$

where λ is a standardized factor loading for item and N is the number of items in a factor.

Since, data related to all three variables of the study were collected at the same time and from same respondents, it may lead to CMV. CMV created serious challenges for a researcher. It may result into prejudiced validity and reliability estimates, exaggerated or flattened correlation coefficients and deflated interaction effect (Podsakoff *et al.*, 2012). In this study, both procedural and statistical remedies were relied for minimizing the proposed variance. First, pre-validated questionnaires having different rating scale for gratitude (seven-point) and leadership (five-point) were used. Second, questionnaire did not recorded names of faculty members to ensure their anonymity. Also, respondents were assured for confidentiality of records and academic use of collected information. These small steps elicit better and true responses from respondents (Podsakoff *et al.*, 2012). Apart from these procedural methods, Harman's one-factor test was also used to check common method bias. More than one factor was reported when items of different measures were subjected to factor analysis. Extraction of more than one factors nullified chances of common method biasness in this study.

Statistical tools

Correlation matrix was generated to explore the relationship between gratitude, SI and leadership. Correlation provides information about degree of association between two variables only. Causal linkage could be determined with the help of linear regression. Thus, causal linkage between gratitude and leadership was investigated using regression. Also, proposed mediation role of SI was examined with the help of Baron and Kenny's (1986) four step technique and bootstrapping method.

Variables	Mean	SD	α	CR	AVE	*G*	SI	LF
G	4.66	1.171	0.798	0.772	0.685	1		
SI	3.53	0.528	0.839	0.812	0.731	0.371**	1	
L	3.61	0.668	0.908	0.895	0.742	0.446**	0.668**	1

Notes: G, gratitude; SI, social intelligence; L, leadership; α, Cronbach's α; CR, composite reliability; AVE, average variance explained. **Significant at the 0.01 level (two-tailed)
Source: Primary data

Table I. Descriptive statistics, reliability, validity, CMV and correlation matrix

Data interpretation and analysis

Table I described descriptive statistics, reliability, validity, CMV statistics and correlation among gratitude, SI and leadership. Findings concluded significantly positive correlation among gratitude and SI, gratitude and leadership and SI and leadership. According to Cohen *et al.* (2003), correlation value falling within $0.1 < r < 0.3$ should be considered as a small effect, value within $0.3 < r < 0.5$ as a medium effect and value within $r > 0.5$ as a large effect. From Table II, we may conclude a medium correlation between gratitude and leadership ($r = 0.446$), and between gratitude and SI ($r = 0.371$), and a higher correlation between SI and leadership ($r = 0.668$). It provides empirical evidence in support of *H1–H3*. Correlation symbolizes degree of association between two variables; however, researchers recommend the use of regression to examine the direct relationship. Accordingly, a regression model – $L = \beta_0 + \beta_1\ \mathrm{GQ} + e$ – was proposed which considers leadership as dependent variable and gratitude as an independent variable.

Tables II and III represented results of proposed regression model. The results showed that gratitude among teachers could explain variation in their leadership ($\beta = 0.446$, p-value < 0.05). Gratitude explained and accounted for as much as about 20 percent variation in a teacher's leadership. Hence, accordingly following regression model is extracted:

$$\mathrm{LF} = 2.424 + 0.446\ \mathrm{GQ}.$$

Now, mediating effect of SI was explored through multiple hierarchical regression, which was applied on following four steps of Baron and Kenny's (1986) mediation analysis:

(1) The predictor or independent variable (gratitude) significantly predicts outcome or dependent variable, i.e. leadership.

(2) The independent variable significantly predicts the mediating variable (SI).

(3) The mediating variable (SI) significantly predicts the outcome variable (leadership).

(4) When the effect of mediating variable is controlled, the association between independent variable and dependent variable no longer remains significant (full mediation) or their association reduces significantly (partial mediation). In case of partial mediation, the regression coefficient reduces on controlling for the mediating variable (Table IV).

Table V elaborated the mediating effect of SI on the relationship between gratitude and leadership. First of all, gratitude positively and significantly influences leadership

Model	Unstandardized coefficient		Standardized coefficients		
	B	SE	*β*	*t*-value	Sig.
Constant	2.424	0.202	–	11.997	0.000*
G	0.254	0.042	0.446	6.005	0.000*

Note: *Significant at the 0.05 level (two-tailed)
Source: Primary data

Table II. Result of linear regression (IV – gratitude, DV – leadership)

Model	*R*	R^2	Adjusted R^2	SE	*F*	Sig.
1	0.446	0.199	0.193	0.600	36.668	0.000*

Note: *Significant at the 0.05 level (two-tailed)
Source: Primary data

Table III. Model summary (IV – gratitude, DV – leadership)

($\beta = 0.446$, $p < 0.05$). Next, we see that gratitude also positively and significantly predicts SI ($\beta = 0.371$, $p < 0.05$). The third condition of Baron and Kenny's mediation process can be seen and met as SI positively influences leadership ($\beta = 0.668$, $p < 0.05$). Finally, after controlling for the SI, it was observed that gratitude still positively and significantly influences leadership, although this impact has reduced from $\beta = 0.446$ to $\beta = 0.229$, $p < 0.05$, thereby fulfilling the last condition and showing a partial mediation. The results concluded that SI partially mediates the relationship between gratitude and leadership among Indian university teachers.

Although Baron and Kenny's (1986) method of mediation provides fair knowledge of mediation effect, yet researchers argue that this method is essentially a parametric test, i.e. it assumes normal distribution of data. Hence, researchers suggest bootstrapping method is based on sampling with replacement done for a large number of times, usually around 5,000 times. This popular technique of checking mediation has been incorporated in the present study as well. Table V shows mediation analysis with the help of bootstrapping estimates. Specifically, indirect effect in bootstrapping table confirmed the presence of mediating effect of SI. In nut shell, both methods (Baron and Kenny's, 1986 method and bootstrapping technique) concluded partial mediating effect of SI amid relationship between gratitude and leadership. Thus, *H4* is accepted here.

Discussion and conclusion

Today, a teacher is not restricted to classrooms only; rather they are supposed to take greater responsibilities in administration, policy making and bringing changes in traditional teaching-learning processes and practices. This unprecedented extension of their scope of work has fascinated the attention of researchers from several areas like psychology, sociology, public administration, organizational behavior, etc. Berg and Zoellick (2019) reported an exponential growth in the research and literature on teacher leadership. Importance bestowed to teacher leadership could be realized from the fact that "American Educational Research Association" has initiated an annual program entitled "Teacher Leadership Congress." This annual congress meeting highlighted the dire need to develop teacher leaders for the betterment of education. In the same pursuit, this study intended to

Table IV.
Mediation analysis (results of hierarchical regression)

Relationship	Standardized coefficients(β)	*t*-value	Sig.	Result
G → L	0.446	6.055	0.000	Significant
G → SI	0.371	4.866	0.000	Significant
SI → L	0.668	10.910	0.000	Significant
G → SI → L	0.229	3.618	0.000	Significant

Notes: G, gratitude; SI, social intelligence; L, leadership
Source: Primary data

Table V.
Mediation analysis (results of bootstrapping technique)

Relationship	Effect	SE	LLCI	ULCI	Result
G → L	0.2543	0.0420	0.1713	0.3373	Significant
G → SI	0.1677	0.0345	0.0996	0.2358	Significant
SI → L	0.7362	0.0801	0.5779	0.8945	Significant
G → SI → L	0.1308	0.0362	0.0594	0.2023	Significant
Indirect effect	0.1234	0.0423	0.0429	0.2082	Significant

Notes: G, gratitude; SI, social intelligence; L, leadership
Source: Primary data

explore the relationship between gratitude and leadership among university teachers. The study reported significant correlation between two variables. It was also found that gratitude explained 20 percent variation in teacher leadership. Also, SI proved to be a partial mediator amid relationship between gratitude and teacher leadership.

These results have significant effects on the efforts of government, private players in education industry and other stakeholders, who are devoting huge amounts of money, time, energy and other resources for the development of leadership skills among teachers. This study proposes gratitude as an antecedent for the same. Howells (2013) emphasized in his study that gratitude as a practice enhances teacher–student relationship and builds positive classroom and learning environment as the broadening and building effect of gratitude leads to a nurturing environment with increased teacher–student bonding and connection. Hence, in this study, it was proposed that gratitude among teachers will enhance their roles as leaders since the spillover effect of gratitude will lead to pro-social reciprocity among students and they would relatively feel more motivated, relaxed and positive under such leadership. The findings of this study are consistent with Howells' (2013) study as it was found that gratitude positively influenced leadership among teachers. Nieuwhof (2017) discussed five crucial motives for leaders to practice and cherish gratitude. Gratitude makes a person more compassionate, kinder, less suspicious and less resentful. And a good leader is supposed to possess these essential human values. Second, gratitude enriches people with sense of abundance. And this abundance mentality allows leaders to see and utilize an opportunity where other misses. Sense of abundance enables leaders to appreciate enormous prospects of growth and development. Third, generosity is a natural outcome of gratitude and it wins hearts and souls of the followers. Fourth, gratitude counterbalances anger and jealousy. And a calm and compose leader can easily touch the skies of success. At last, gratitude possesses power of attraction. Researchers revealed that people are to accept lower wages or even pay cuts to work under a grateful leader. The findings of the present study complement with these results of Nieuwhof (2017).

This study also concluded that SI partially mediated this positive relationship between gratitude and leadership. Through the broadening mechanism, grateful teachers were more able to understand and infer their students' thoughts and empathize with them, they expressed sensitivity to students' needs, therefore building stronger bonds as well as enhanced skills with a broader outlook and hence could provide more of inspiration, motivation, intellectual stimulation and individual consideration – all of which are important qualities of an effective leader. Thus, Fredrickson's (2004) broaden-and-build theory of positive emotion gets another explanation through this study. Gratitude (positive emotion) broadens SI through influencing social awareness level, social information processing and social skills, which builds leadership skills among gratitude practitioner.

These findings have major practical implications with regard to higher educational settings. It is important that teachers, who are the promoters of knowledge and change, must themselves operate in a gratuitous environment. This is consistent with previous studies that have focused on cultivating employee gratitude in organizations. It is important that higher institutions promote institutionalized gratitude (gratitude shared by all teachers) with expressions of gratitude starting from the top management itself, thereby creating a trickle-down effect as emotions are contagious. Teachers must be thanked and appreciated by the higher authorities for their tremendous efforts and hard work that they put in through the students' educational journey. The appreciation must involve intrinsic as well as extrinsic rewards, hence instilling gratefulness in teachers, thereby encouraging them to put in extra efforts and to become better role models to their students. It is suggested that educational institutions promote several ways to cultivate gratitude. These ways include gratitude practices and interventions including gratitude journals, gratitude letters, gratitude visits, mindfulness, gratitude collages and sharing positive events. Waters and

Stokes (2015) found effective influence of "gratitude diary" in triggering emotion gratitude and "gratitude letter" in triggering action gratitude among school leaders. These interventions could be practiced easily at workplace through seminars, courses, conferences and workshops. Some new innovative gratitude practices have emerged such as "haiku," poem of three lines expressing gratefulness and "Naikan therapy" which is a popular Japanese technique of practicing gratitude in organizations. Meditation and progressive muscle relaxation have also been cited as techniques to induce gratitude (Shapiro *et al.*, 2002; Khasky and Smith, 1999; Emmons and Shelton, 2002). Emmons and Shelton (2002) emphasized on fostering a grateful culture, demanding that gratitude be cultivated as a virtue in organizations rather than as a transient emotion.

The present study provides a mediating model of gratitude, SI and leadership; however, these contributions come in the light of a few limitations. First, data are collected using self-reported measures which may have caused common method biasness. As recommended by Podsakoff *et al.* (2012), established validated scales were used and also anonymity of respondents was strictly observed to reduce CMV. Future researchers may use multiple sources of data collection or may resort to longitudinal research design. Second, the study is based on a sample of 150 teachers of Delhi-NCR only. A larger sample comprising of respondents from diverse cities would be appropriate for generalization of the results in Indian setting. Future research should explore a larger sample spread over different cities across India. Third, this study explored only the mediating effect of SI. Future researchers may even explore effects of other possible mediators in the relationship between gratitude and leadership such as teacher work attitudes (job satisfaction, job involvement and organizational commitment), motivation, self-esteem and perceived organizational support. And at last, future research may investigate the role of personality traits, stress, power and conflict, politics and perceived organizational justice as possible moderators in the relationship between gratitude and leadership.

References

Algoe, S.B., Haidt, J. and Gable, S.L. (2008), "Beyond reciprocity: gratitude and relationships in everyday life", *Emotion*, Vol. 8 No. 3, pp. 425-429.

Avolio, B., Walumbwa, F. and Weber, T.J. (2009), "Leadership: current theories, research, and future directions", *Annual Review of Psychology*, Vol. 60 No. 1, pp. 421-449.

Baron, R.M. and Kenny, D.A. (1986), "The moderator-mediator variable distinction in social psychological research – conceptual, strategic, and statistical considerations", *Journal of Personality and Social Psychology*, Vol. 51 No. 6, pp. 1173-1182.

Bass, B.M. (1990), "From transactional to transformational leadership: learning to share the vision", *Organizational Dynamics*, Vol. 18 No. 3, pp. 19-31.

Bass, B.M. and Avolio, B.J. (1992), "Developing transformational leadership: 1992 and beyond", *Journal of European Industrial Training*, Vol. 14 No. 5, pp. 21-27.

Berg, J.H. and Zoellick, B. (2019), "Teacher leadership: toward a new conceptual framework", *Journal of Professional Capital and Community*, Vol. 14 No. 1, pp. 2-14.

Burns, J.M. (1978), *Leadership*, Harper & Row, New York, NY.

Cohen, J., Cohen, P., West, S.G. and Aiken, L.S. (2003), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd ed., Erlbaum, Mahwah, NJ.

Cronbach, L.J. and Shavelson, R.J. (2004), "My current thoughts on coefficient alpha and successor procedures", *Educational and Psychological Measurement*, Vol. 64 No. 3, pp. 391-418.

Crowther, F., Ferguson, M. and Hann, L. (2009), *Developing Teacher Leaders: How Teacher Leadership Enhances School Success*, Corwin Press, Thousand Oaks, CA.

Emmons, R.A. and McCullough, M.E. (2003), "Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life", *Journal of Personality and Social Psychology*, Vol. 84 No. 2, pp. 377-389.

Emmons, R.A. and Shelton, C.M. (2002), "Gratitude and the science of positive psychology", *Handbook of Positive Psychology*, Vol. 18 No. 1, pp. 459-471.

Eshghi, P., Etemadi, M., Mardani, M., Fanaei, E. and Agha-hosaini, T. (2013), "Social intelligence and its sub-scales among physical education expertise in Isfahan education organizations: study of gender differences", *European Journal of Experimental Biology*, Vol. 3 No. 4, pp. 13-17.

Fornell, C. and Larcker, D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18 No. 1, pp. 39-50.

Fredrickson, B.L. (2004), "The broaden-and-build theory of positive emotions", *Philosophical Transactions of the Royal Society B: Biological Sciences*, Vol. 359 No. 1449, pp. 1367-1378.

Froh, J.J., Emmons, R.A., Card, N.A., Bono, G. and Wilson, J.A. (2011), "Gratitude and the reduced costs of materialism in adolescents", *Journal of Happiness Studies*, Vol. 12 No. 1, pp. 289-302.

Garg, N. (2017b), "Workplace spirituality and organizational commitment: an empirical exploration", *Purushartha – A Journal of Management, Ethics and Spirituality*, Vol. 10 No. 2, pp. 8-24.

Garg, N. (2017a), "Workplace spirituality and employee well-being: an empirical exploration", *Journal of Human Values*, Vol. 23 No. 2, pp. 1-19.

Garg, N. (2017c), "Workplace spirituality and organizational performance in Indian context: mediating effect of organizational commitment, work motivation and employee engagement", *South Asian Journal of Human Resources Management*, Vol. 4 No. 2, pp. 191-211.

Garg, N. (2018), "Promoting organizational performance in Indian insurance industry: the roles of workplace spirituality and organizational citizenship behaviour", *Global Business Review*, 0972150918778983.

Garg, N., Punia, B.K. and Jain, A. (2019), "Workplace spirituality and job satisfaction: exploring mediating effect of organization citizenship behaviour", *Vision: The Journal of Business Perspective*, Vol. 23 No. 3, pp. 287-296.

George, D. and Mallery, P. (2003), *SPSS for Windows Step by Step: A Simple Guide and Reference 11.0 Update*, 4th ed., Allyn & Bacon, Boston, MA.

Goleman, D. and Boyatzis, R. (2008), "Social intelligence and the biology of leadership", *Harvard Business Review*, Vol. 86 No. 9, pp. 74-81.

Howells, K. (2013), "The role of gratitude in helping school leaders strengthen relationships", *Perspectives on Educational Leadership*, Vol. 6 No. 1, pp. 1-3.

Khasky, A.D. and Smith, J.C. (1999), "Stress, relaxation states and creativity", *Perceptual and Motor Skills*, Vol. 88 No. 1, pp. 409-416.

McCullough, M.E., Kilpatrick, S., Emmons, R. and Larson, D.B. (2001), "Is gratitude a moral affect?", *Psychological Bulletin*, Vol. 127 No. 2, pp. 249-266.

Mangin, M.M. and Stoelinga, S.R. (2008), *Effective Teacher Leadership: Using Research to Inform and Reform*, Teachers College Press, New York, NY.

Marlowe, H.A. (1986), "Social intelligence: evidence for multidimensionality and construct independence", *Journal of Educational Psychology*, Vol. 78 No. 1, pp. 52-58.

Neumerski, C.M. (2013), "Rethinking instructional leadership, a review: what do we know about principal, teacher, and coach instructional leadership, and where should we go from here?", *Educational Administration Quarterly*, Vol. 49 No. 2, pp. 310-347.

Nieuwhof, C. (2017), "5 Crucial ways your gratitude impacts your attitude as leader", available at: https://careynieuwhof.com/5-ways-your-gratitude-impacts-your-attitude-as-a-leader/ (accessed November 12, 2019).

Podsakoff, P.M., MacKenzie, S.B. and Podsakoff, N.P. (2012), "Sources of method bias in social science research and recommendations on how to control it", *Annual Review of Psychology*, Vol. 63 No. 1, pp. 539-569.

Ritzenhöfer, L., Brosi, P., Spörrle, M. and Welpe, I. (2017), "Leader pride and gratitude differentially impact follower trust", *Journal of Managerial Psychology*, Vol. 32 No. 6, pp. 445-459.

Shapiro, S.L., Schwartz, G.E.R. and Santerre, C. (2002), "Meditation and positive psychology", in Snyder, C.R. and Lopez, S. (Eds), *Handbook of Positive Psychology*, Oxford University Press, London, pp. 632-645.

Silva, D.Y., Gimbert, B. and Nolan, J. (2000), "Sliding the doors: locking and unlocking possibilities for teacher leadership", *Teachers College Record*, Vol. 102 No. 4, pp. 779-804.

Silvera, D.H., Martinussen, M. and Dahl, T.I. (2001a), "The Tromsø Social Intelligence Scale, a self-report measure of social intelligence", *Scandinavian Journal of Psychology*, Vol. 42 No. 1, pp. 313-319.

Silvera, D.H., Martinussen, M. and Dahl, T.I. (2001b), "The Tromsø Social Intelligence Scale, a self-report measure of social intelligence", *Scandinavian Journal of Psychology*, Vol. 42 No. 4, pp. 313-319.

Spillane, J.P. (2006), "Distributed leadership", *The Educational Forum*, Vol. 69 No. 2, pp. 143-150.

Waters, L. and Stokes, H. (2015), "Positive education for school leaders: exploring the effects of emotion-gratitude and action-gratitude", *Australian Educational and Developmental Psychologist*, Vol. 32 No. 1, pp. 1-22.

Wood, A.M., Froh, J.J. and Geraghty, A.W. (2010), "Gratitude and well-being: a review and theoretical integration", *Clinical Psychology Review*, Vol. 30 No. 7, pp. 890-905.

Wood, A.M., Joseph, S. and Maltby, J. (2008), "Gratitude uniquely predicts satisfaction with life: Incremental validity above the domains and facets of the five factor model", *Personality and Individual Differences*, Vol. 45 No. 1, pp. 49-54.

Further reading

Malhotra, N.K. and Dash, S. (2011), *Marketing Research an Applied Orientation*, Pearson Publishing, London.

Corresponding author
Naval Garg can be contacted at: naval.garg@dtu.ac.in

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

Check for updates

Original Article

How Can Small and Medium Enterprises Effectively Implement Corporate Social Responsibility?: An Indian Perspective

Global Business Review
1–29
© 2019 IMI
Reprints and permissions:
in.sagepub.com/journals-permissions-india
DOI: 10.1177/0972150919865086
journals.sagepub.com/home/gbr
SAGE

Gaurav Dawar[1]
Seema Singh[2]

Abstract

The current study is a strategic approach to corporate social responsibility (CSR); the aim is to put forward the factors of CSR activities that enhance its effectiveness for small and medium enterprises (SMEs). To achieve this objective, the factors were extracted from the literature and described along with trusteeship theory of Mahatma Gandhi, and an exploratory study was conducted and data were collected using structured questionnaire based on pretested scale from 158 SMEs and tested using partial least square regression (PLSR). The statistics shows the overall model fit, and the findings indicate a significant relationship with effective CSR. The results of the study are in accordance with the previous research work, and we also find that environment-related CSR and partnership are crucial for the effectiveness of CSR in SMEs, stakeholders' role are important and SMEs' CSR practice is still informal. The variables identified from study will help SMEs in establishing a formal approach towards CSR and meeting the needs of business and society in the twenty-first century.

Keywords

Effective corporate social responsibility, small and medium enterprises, partial least square regression, trusteeship theory

Introduction

Integration of economies due to globalization has brought tremendous changes in the nature and structure of market as well as the expectation of stakeholders from organizations. With the increase in social awareness in the last decade (De Clerecq & Voronov, 2011), companies are now facing pressure to meet the expectations of market participants, other than shareholders and investors, seeking multiple

[1] Assistant Professor, IMS Ghaziabad, UP, India.
[2] Professor, Department of Humanities, Delhi Technological University, Delhi, India.

Corresponding author:
Gaurav Dawar, Assistant Professor, IMS Ghaziabad, UP - 201009, India.
E-mail: gauravdawar@gmail.com

objectives. It implies that market participants not only affect the traditionalistic aim of profit maximization but also through their choices, directly and indirectly, affect and contribute to society. The large corporations in India are largely engaged in corporate social responsibility (CSR) so as to enhance their reputation with key stakeholders and are better resourced at it. However, for small and medium enterprises (SMEs), engagement in CSR is based on reasons, expected benefits and the ways of practising CSR (Smith, 2013; Jamali et. al., 2009). The adoptions of CSR activities for SMEs are exposed to barriers such as limited resources and capabilities possessed by them. It also leads to another concern of identifying and prioritizing stakeholders for designing CSR policy (Brammer & Millington, 2003). The solution to these challenges is to focus on enhancing the effectiveness of CSR activities that can derive the value for SME (López-Pérez, Melero, & Javier Sese, 2017).

CSR is a tool leveraged by the organizations to fulfil societal expectation and meet regulatory obligations (Balakrishnan, Malhotra, & Falkenberg, 2017; Carroll, Lipartito, Post, & Werhane, 2012; Freeman & Velamuri, 2006; Goodpaster, 2013; Margolis & Walsh, 2003; Dawar and Singh, 2016). The discipline of CSR has long been studied in developed nations (Carroll & Buchholtz, 2012; Crane, Matten, & Spence, 2014) and now the attention is diverted to emerging economies to further contribute to knowledge and practices of social responsibility which is considered as a umbrella term to describe the multifaceted relations of business with society to account for economic, social and environmental impact of business activities in developing nations (Jamali, Lund-Thomsen, & Jeppesen, 2017). A quantum of research has been carried out on CSR in the twenty-first century that has made it a global norm for business, yet the debate continues as to what contributes to the effectiveness of CSR.

SMEs are important backbone of healthy economy in India because they account for 60 per cent of total manufacturing employment and 40 per cent of national income (Beck, Demirguc-Kunt, & Martinez Peria, 2008) and are closely linked to local community in developing countries. Yet they have received relatively less attention in the broad research agenda pertaining to the role of SME in CSR in emerging nations (Harris, Sapienza, & Bowie, 2009; Jamali et al., 2017). SMEs follow informal approach to CSR (Baumann-Pauly et al., 2013; Jenkins, 2004) due to high constraints of resources and capabilities. Though the relationship between financial performance and CSR (Jain, Vyas, & Chalasani, 2016) has been explored in a developing nation, the area demands substantial attention. Hence, moving further with firm-level analysis, it is also important to affirm a grounded understanding of how CSR practices can be made effective considering institutional and contextual factors along with the pressure of global sustainability. Within this context, this article aims to investigate the variables that lead to the effectiveness of CSR activities in SMEs in India.

Effective from April 2014, the Indian government mandated CSR spending under companies act to be followed by both private- and public-sector companies that fulfil its criteria of net worth of ₹50 million or more, turnover of ₹100 million or more or net profit of ₹0.5 million. These guidelines follow distinctively Indian approach deriving its existence from the trusteeship model propounded by Mahatma Gandhi. Trusteeship is an inclusive model that aims to transform the capitalist order of the society into an egalitarian one by asserting the obligation of the wealthy to use their wealth for the benefit of underprivileged. As a results of significant developments in the Indian Economy over the years four approaches of CSR emerged, that is, trusteeship, statist model, liberal model and stakeholder model (Sharma, 2009) and of these the view of Mohandas Karamchand Gandhi on corporate responsibility is influential in Indian corporate sector (Balasubramanian, 2010). Gandhiji's theory of trusteeship focusses on consolidation and amplifying social development as he regarded Indian companies as 'Temples of Modern India' as these industrial houses build trust to enhance social reforms of rural development, women empowerment and education. Gandhi's thoughts have already been successfully applied in the

area of leadership (Axelrod, 2010) and corporate responsibility (Balakrishna et al., 2017) and considering their global relevance, they merit further academic attention.

In this respect, a critical question is posed to the researcher that 'can trusteeship theory of Mahatma Gandhi help in identifying the factors that lead to effective corporate social responsibility (ECSR)'?

Building on trusteeship theory, we propose an ECSR approach to be adopted by SMEs. The suggested framework can enhance the effectiveness of CSR activities along with its strategy in practice and contribution to sustainability. Further, we test the framework by using Partial least square structure equation modelling with data from 158 SMEs in India. In doing so, this study makes several contributions to theory and practice. As first contribution, this study advances the debate on CSR adoption in SMEs and its value by proposing an SME-specific approach of CSR. As second contribution, this study identifies and discusses the factors that improve effectiveness of CSR activities for SMEs. As third contribution, this study puts forward several implications for managers and policymakers.

The remainder of the article is structured as follows: the second section provides a discussion on the background of the study and highlights gaps that this article addresses, the third section includes the methodological aspects of this study, the fourth section discusses the results of this study, the fifth section presents the outline of research implications and the sixth section includes conclusion, limitations and future research directions.

Background of Study and Proposed Hypothesis

CSR has emerged as an essential conceit in the business today. The concept conceived by Howard Bowen in his book *Social Responsibilities of the Businessman* in 1953 has marked a foundation stone for exploration in the area of CSR. Through this book, Bowen wanted to point out towards the *Obligations of the businessman to pursue those policies, decisions and line of actions that are desirable regarding the objectives and values of our Society*. The definition given by Ethics in Action Award (2003):

> CSR is the term describing a company's obligation to be accountable to all of its stakeholders in all its operations and activities, Socially Responsible companies cover full scope of their impact on communities and the environment when making decisions, balancing the needs of stakeholders with their decision to make profit.

The trusteeship theory proposed by Mahatma Gandhi defined the social responsibility of industries and focussed on growth with equity to achieve an inclusive society. These views highlight the importance of shared values between firm and society, and also recognize that the CSR procedure may stay informal for most SMEs. The current study adopts the views of Mahatma Gandhi through trusteeship theory as it is uniquely Indian. The review of the theoretical approaches to CSR and justification of this approach through the framework of ECSR for SMEs are presented in the following section.

Theoretical Approaches to Corporate Social Responsibility

The term CSR has a broader implication, and it encompasses the concept of corporate philanthropy, corporate citizenship and organizational social performance (Feil, 2012). The extant of literature brings out theoretical aspects of CSR as defined by authors like Frederick (1960), Elibert and Parket (1973), Caroll (1979, 1983) and Hopkins (1998); and as a part of theories such as Shareholder theory (Friedman, 1970), Stakeholder theory (Freeman, 1984) and Stewardship theory (Donaldson & Davis, 1991). These

approaches share some limitations: prioritization problem in identification of stakeholders that can influence the firms' performance (Brammer & Pavelin, 2004), failure to identify CSR activities that can enhance firm value and a limited implication to practise (Porter & Kramer, 2006; Schenider & Barsoux, 2003). Although each of these approaches has its own particular merits, the confinements of these methodologies frequently require a combination of theoretical lens when examining CSR and its suggestions for firms (Mcwilliams, Siegel, & Wright, 2006).

Corporate Social Responsibility: Indian Perspective

India is a land of paternalistic philanthropy (Gautam & Singh, 2010; Prabhakar & Mishra, 2013). Chanakya, a well-known economist, and Mahatma Gandhi, a social reformer, have preached the ethical principles of doing business (Sharma & Tewari, 2017). The CSR outlook in India is unique because of 'Family-centred' style of management, and most of the firms are controlled by family groups (Sundar, 2000).

Table 1. Review of Effective Corporate Social Responsibility

Sr. No.	Author (year)	Effective Corporate Social Responsibility Defined	Variables	Research Unit	Industry	Country	Reference Framework
1	Alvarado-Herrera, Bigne, Aldas-Manzano, and Curras-Perez (2017)	CSR scale (CSRCons PerScale) to measure the effectiveness of the CSR programme from sustainable development (SD) perspective from a consumer perspective.	Social, environmental and economic	Consumer (stakeholder perspective)	Tourism	24 countries	Caroll's pyramid and SD
2	Jaiyeoba, Adewale, and Quadry (2018)	firm's activity is to maximize and create value for all the stakeholders	investment development, human development, poverty reduction and environmental development	customers, local communities, employees, and branch managers (stakeholder perspective)	Banking	Malaysia	Caroll's (2006) approach

(Table 1 Continued)

(Table 1 Continued)

Sr. No.	Author (year)	Effective Corporate Social Responsibility Defined	Variables	Research Unit	Industry	Country	Reference Framework
3	Chebet and Muturi (2018)	Formulating strategies to ensure effective corporate social Responsibility to enhance organizational performance	Philanthropic activities, ethical activities, economic activities, legal activities	Sony and Chemelil Sugar Factories, Kenya	Sugar Industry	Kenya	
4	Krisnawati, Yudoko, and Bangun (2018)	CSR can be made effective by involvement with community and compliance with regulations and regularly assessing the CSR activities through feedback evaluation	Community and Regulations (IDV) Output of CSR (DV) CSR planning, organizing, executing and controlling (mediating variables)	Employees of WIKA (construction company of Indonesia)	Construction	Indonesia	
5	Tugrul and Lee (2017)	Build consumer trust, encourage followers to engage, participate, and become active	Message goal frame and message orientation frame	Technical and administrative staff of university	Telecommunication (Turkcell) and NGO (CYDD)	Turkey	Construal-level theory (CLT)
6	Wood (2017)	A strategic business focus supporting shared value to achieve the outcomes from strategic integration of competitive advantage and responsive differentiation to meet social contract through effective corporate social responsibility	Strategic business focus, Competitive advantage, CSR shared value, organizational core values, talent and mindset				Fish, A. J. (2015) 'Shared Value: Reforming and Reframing Strategic Business Focus through Strategy Interdependence, Cultural Complementarity and Balance and Soul'; *White Paper*, *Centre for Progressive Leadership, London Metropolitan University*, ISSN—2396-8591

(Table 1 Continued)

(Table 1 Continued)

Sr. No.	Author (year)	Effective Corporate Social Responsibility Defined	Variables	Research Unit	Industry	Country	Reference Framework
7	Adeneye and Ahmed (2015)	Shareholders wealth is maximized through effective corporate social responsibility practices	CSR and company's financial performance through return on capital employed, market book value and firm size	500 UK firms	All	UK	
8	Sharma and Tewari (2017)	HR leadership and employee involvement are central to the effective institutionalization of CSR	Organizational culture, corporate capability, perception towards CSR activities and perceived corporate performance	180 employees and HR professionals of private and public organizations	Private sector—financial institutions, IT solution providers, foods and beverages, petrochemical and telecom industry, and three public sector companies	India	
9	Nwoke (2017)	CSR contribution to social and economic development. effective CSR is the social responsibility obligation that contributes to SD	Ideological, practical and political barrier				Neo-liberalism and shareholder value
10	Collier and Esteban (2007)	The effective delivery of corporate social and environmental responsibility initiatives are dependent on employee responsiveness	Employee involvement, employee commitment to towards organizational justice and fairness and priority of top management			USA and UK	

(Table 1 Continued)

(Table 1 Continued)

Sr. No.	Author (year)	Effective Corporate Social Responsibility Defined	Variables	Research Unit	Industry	Country	Reference Framework
11	Vlachos, Panagopoulos, Bachrach, and Morgeson (2017)	The effectiveness of CSR has linkage with employee motivation	CSR casual attribution, employee advocacy, manager's organizational tenure	Employees of the manufacturing organization in Europe	Energy and automation technologies; aluminium solutions; and cosmetics and pharmaceutical products	Europe	Ferris et al. (1995) framework: employees seek to make sense of other employees' prosocial behaviours, they develop attributions of intentionality
12	Batool, Butt, and Niazi (2016)	Adapting to CSR practices is helping companies keep hold of their customers, who are much more aware and knowledgeable, regarding their rights and responsibilities towards the environment	CSR practices, CSR practices rating, CSR effectiveness	KSE 100, SMEs, NGOs and service providers	Manufacturing companies, telecom companies, tobacco companies and chemical companies	Pakistan	
13	Kim, Cho, Kim, and Lee (2011)	Philanthropic activities of CSR dimensions contribute positively towards effectiveness	Who, what and for whom	Students and consumers of different age groups	All		Caroll's (1979) approach
14	Rim and Song (2013)	Communication of CSR practices at different platform can make CSR effective	Prior corporate reputation; corporate blog responsiveness; interaction between previous reputation and blog responsiveness	Undergraduate students of university			

(Table 1 Continued)

(Table 1 Continued)

Sr. No.	Author (year)	Effective Corporate Social Responsibility Defined	Variables	Research Unit	Industry	Country	Reference Framework
15	Murray and Vogel (1997)	CSR effectiveness is related to addressing the stakeholder expectations	The occupational setting, voting behaviour, political and social behaviour, media habits	Managers		USA	Hierarchical perspective

Source: The authors.

Note: IDV - Independent Variable, DV - Dependent Variables, KSE - Karachi Stock Exchange.

The selections of social responsibility activities are influenced by cultural and social preferences of the family (Kansal et al., 2014). India in the twenty-first century needs a holistic approach that can bridge the gap between the different socio-economic sections of society, not always requiring state action, to achieve sustainable future, thus trusteeship-based framework is suitable for value creation. Gandhiji believed that trusteeship theory would lead to sustainable economic development (Gopinath, 2005; Balakrishnan et al., 2017) because it blends technology, environment, people, economics and culture. The approach may also enhance the role of affluent entrepreneurs in nation building (Rolnick, 1962).

The philosophy of trusteeship as defined by Mahatma Gandhi:

> Supposing I have come by a fair amount of wealth—either by way of legacy, or by means of trade and industry—I must know that all that wealth does not belong to me; what belongs to me is the right to an honorable livelihood, no better than that enjoyed by millions of others. The rest of my wealth belongs to the community and must be used for the welfare of the community.
> —Mahatma Gandhi,Theory of Trusteeship, www.mkgandhi.org

Gandhiji advocated the application of trusteeship at multiple levels to attain equality in society. The theory emphasizes upon the intelligent use of wealth, believes in the strength of considering others equally and motivates to respect everyone's rights, priority towards of commitment to principles of social reformation rather than policies and affluent people should part their wealth towards local schemes and plans. This theory is a feasible solution to the social issues prevalent at the beginning of the previous century and predictable future (Prasad, 2013). The concepts of trusteeship have been incorporated into the CSR legislation implemented in the year 2014, so that a significant number of companies contribute in nation building. The firms that come under the new act should form a CSR committee from existing board members, devise CSR policy, and implement and monitor CSR activities. Rahul Bajaj, Chairman Bajaj Auto Ltd., emphasized on the adoption of a modified version of trusteeship model for meeting the contemporary needs of business and society in the twenty-first century. The CSR models derived from trusteeship theory involves practising the social responsibility alone or in partnership with other organizations or in association with community or combination of all, but the essence of CSR can be maintained from the vision of board, performance of responsible leaders, following the environmental regulations, creating social and economic value through every aspect of business (Bajaj, 2016).

At present scenario, CSR is required to be innovative concerning processes and new management thinking to make it effective. The concept of ECSR is discussed in the next section.

Effective Corporate Social Responsibility

As per the online dictionary *Merriam Webster*, Effective means *producing a result that is wanted.* Firm managing resources effectively generate capabilities that are valuable, rare, imitable and non-substitutable (Barney, 1991). ECSR involves giving back to the society by fulfilling obligations to various stakeholders of organization (Szczanowicz & Saniuk, 2014). It attracts tangible and intangible benefits like SD, access to skilled human capital, reduction of cost and increase in the innovation quotient (Lydenberg, Rogers, & Wood, 2010; Szczanowicz & Saniuk, 2014; Babatunde & Adeyemi, 2015). The concept was first discussed by Pavel Castka (2004) and emphasized *organizations to operate profitably yet in socially and environmentally responsible manner to achieve business sustainability* and after that, various authors have explored it through different frameworks such as Caroll's pyramid and construal level theory, etc. (Table 1). A significant amount of research has been undertaken in countries, both developed and developing, yet further investigation is required to explore the compelling factors of ECSR specially in context to manufacturing SMEs. Also, trusteeship theory has not yet been utilized earlier for identifying the factors that improve the efficiency of CSR practices of the organization. This is where our study makes a significant contribution. The study further identifies variables from literature and describes them along with the thoughts of Gandhiji.

Environmental Regulations and Corporate Social Responsibility

Gandhiji had a deep sense of concern for the natural environment that accompanies industrialization and modernity to maintain ecological balance (Kolge & Sreekumar, 2011). His thoughts and views are not intended towards a formal sustainable environmental structure but interlinking his ideas that can lead to a SD model (Daptardar, 2012) as he expected society and owners to be steward of natural environment.

Supporting environmental improvement is one of the principal reasons for companies to behave environmentally responsibly in their practices (Florida & Davidson, 2001; Morrow & Rondinelli, 2002). Organizations look for ecologically beneficial practices that may cause less impact on the environment and efforts to preserve the natural habitats (Dolsak & Prakash, 2016; Kumar & Polonsky, 2017). The majority of organizations' CSR can relate to the views and thoughts of Gandhiji on environmental regulations. The Literature points out that SMEs, through continual learnings, can effectively achieve the environmental responsibility (Stewart & Gapp, 2014). An effective CSR-based business model can be achieved by undertaking social and environmental impact assessment and communicating the prospective benefits to the community that can be delivered (Slack, 2012). Thus, following environmental regulations certainly makes an impact on the existence of the organization and its CSR policy (Williamson, Lynch-Wood, & Ramsay, 2006).

Hence, the following hypothesis is proposed:

> *H1: Supporting environmental regulations significantly contributes to effective corporate social responsibility.*

Integration of Corporate Social Responsibility into Corporate Policy

Corporate scandals by big companies like Enron in the international context and Satyam in the Indian context have created the need of the business to conduct themselves to make their existence legitimate, socially rational and genuine (Rassindren & Hans, 2015). The Indian social responsibility theory of trusteeship proposed by Mahatma Gandhi exclaims that managers in the businesses are responsible for the creation of a prosperous society (Balakrishnan et al., 2017). The underlying idea is to prepare managers as public trustee, and to maintain balance between the corporate resources and support of good causes. It can be achieved by integrating CSR into every activity of the business and making it a part of a policy at the business level. Literature has highlighted that CSR is no longer an option for SMEs instead a key for sustainable business strategy (Policies & Practices, 2012). The International Integrated Reporting Framework (IIRC, 2013) has emphasized upon linking the firm's strategy and CSR as it intends to deliver effective governance and performance for creating long-term and short-term values. In current times, the organizations are pledging to International bodies around the world like United Nations Global Compact (UNGC), International Labour Organisation (ILO), International Organisation for Sustainable Development (IOSD) and European Commission to show the conduct of Global citizenship in the area of labour and human rights, environmental protection and SD (Wharton, 2012), and it can indeed be achieved by integrating CSR into everyday business. Thus, we formulate the following hypothesis:

> *H2: Integration of CSR into corporate policy significantly contributes to effective corporate social responsibility.*

Priority of Board

There are various reasons for heightened Board of Directors' attention towards CSR, that is, moral obligations, sustainability, licence to operate and reputation, and carefully building on it leads to competitive advantage for the organization (Porter & Kramer, 2006). As a result, the Board of Directors and CEOs are devoting higher attention towards CSR and taking it up as a formal structure in the organization (Crespin, 2012). In the recent years, CSR has gained a strategic advantage in the eyes of Board, and it has resulted in the widespread engagement of corporations in various CSR activities including environmentally friendly practices, supporting social causes, financing employees' education and encouragement of ethical training programmes (Maignan & Ferrel, 2000). The trusteeship model has also emphasized the vision on the board of directors as it is important for the essence of CSR (Bajaj, 2016). Cannon (1994) considers directors as responsible for trusteeship of corporate assets primarily to stakeholder. In fact, Gandhiji's trusteeship theory has revolutionary implications as it aims at developing a participative culture and promotes egalitarian values throughout the organization (Ghosh, 1989).

Thus, the following hypothesis is proposed:

> *H3: Board of directors' priority towards social commitment significantly contributes to effective corporate social responsibility.*

Stakeholder Involvement

The term *Stakeholder* involves all the unit of relationship between business, groups and individuals who can affect or are affected by the firm (Freeman et al., 2010). The survival of the organization is infeasible without their support, and hence, social responsibility activities should be customized to their involvement (Freeman, 1984). Stakeholder theory lays its foundation on the concepts of planning, system theory, organizational theory and CSR to develop an integrated framework social responsibility (Freeman & McVea, 2001). Freeman and Velamuri (2008) has also supplemented with the integration of *Company Stakeholder Responsibility* to CSR for further elaboration on previous

developments. The Gandhian trusteeship theory and stakeholder theory share synergies as both lay emphases on working towards the community. Balakrishnan et al. (2017) have highlighted that the Gandhian trusteeship model integrates the role and responsibilities of various stakeholders towards economic value creation and building a prosperous society. His thoughts on equitable system to manage societal resources have been endorsed by government committees (Joseph, Mahodaya, & Pradhan, 2016). The notion is all concerned stakeholders and actors must give a serious thought to manage the societal and natural resources on trusteeship lines, lest the problem assume an endemic form (Joseph et al., 2016). Engaging stakeholders in social responsibility is an essential ingredient for the successful execution of CSR and its reporting (Carroll, 1991; Greenstone, 2014; Heismann, 2014; Russo & Perrini, 2010). Ramus and Steger (2000) have laid on the importance of workers in designing environmental policy to secure an ECSR programme. Also designing an ECSR programme to keep employees motivated is a responsibility of the organization. Thus from the above discussion, the following hypothesis is proposed:

H4: Stakeholder involvement significantly contributes to effective corporate social responsibility.

Responsible Leadership

Effective leadership has a key role in the integration of sustainability practices in the operations of the organization (Adam, 2013). The leaders are responsible for making meaningful transformations that can lead to the creation of socially responsible organizations (Székely & Knirsch, 2005). These executives are accountable to stakeholders, and their responsibility is extended to organizations' societal and environmental influences (Wade, 2006). Thus, responsible leadership is a multilevel solution to emerging ethical, social and environmental challenges (Pless & Maak, 2011). The Gandhian trusteeship theory is closely linked to responsible leadership as it focusses on the moral and socially sustaining responsibility of trustees or leaders towards different stakeholders for economic and social issues. However, Cameron (2011) has enhanced the concept by integrating the component of 'goodness' in all aspects and linking it with 'Virtuous Leadership' that seeks to optimize wealth creation and values for all the stakeholders by creating a relationship of trust and commitment with them (Caldwell, Bischoff, & Karri, 2002; Caldwell, Hayes, Bernal, & Karri, 2008; Hernandez, 2008). The role of Responsible leadership is not limited to ethical conduct (Brown & Treviño, 2006); however, they take an active role in implementing triple bottom line, that is, social, environmental and economic due to stakeholder pressure (Maak, 2007; Pless, 2007). Since it is an integration of the concepts of leadership and social responsibility (Waldman, 2011) and it has a significant impact on responsible business practices of large and medium enterprises (Hansen, 2008), the following hypothesis is proposed:

H5: Responsible leadership significantly contributes to effective corporate social responsibility.

Partnering for Corporate Social Responsibility

Gordic Bader, President of Scott Bader Company in England, has defined trusteeship theory as 'Fraternal partnership between all the factors of production, with the aim of achieving larger social benefit rather than working towards a narrow economic objective such as profit'. Scott Barder Company established in the early twentieth century and implemented the trusteeship principles in 1951 has become one of the successful examples of theory in action (Bader, 1995). But considering the concerns of contemporary economies today the trusteeship institution can be practically implemented through partnership (Kapoor, 2005). The model of an alliance between business organizations with the non-profit organization (NPO) was intended to the creation of knowledge for the development of the voluntary code of conduct for contribution to SD (Arya & Salk, 2006). The Large and medium organizations over the years have

accepted the relevance of non-profit partnership, and it has become a key driver for NPO in fulfilling their overall mission (Álvarez-González, García-Rodríguez, Rey-García, & Sanzo-Perez, 2017; Barross-Mendez, M. J et. al., 2015; Idemudia, 2009). The non-profit partnerships with peers are an established strategy followed by the significant number of business organizations as a part of their broader CSR programme (Mutch & Aitken, 2009). However, despite it being followed at significant organization a little attention has been devoted towards exploration to ECSR. Thus, the following hypothesis is proposed:

H6: Partnering for CSR significantly contributes to effective corporate social responsibility.

Thus, the hypotheses developed from the literature are further tested with the survey of members of the CSR committee at small and medium manufacturing enterprises in the area of National Capital Region. The methodology for testing the hypothesis is discussed in the subsequent section.

Methodology

ECSR is an evolving field, and through the current study, an attempt has been made to investigate the framework for developing nations. Though a significant number of studies were also carried out at developing countries that have used questionnaire items to explore the key variables that may lead to successful implementation of CSR at various industries, Sharma and Tewari (2017) have conceptualized ECSR and its linkage with the involvement of employees. Hence, such approach of exploration of factors in response to questionnaire items seems to be valid for such type of studies. Also, the questionnaire method is the most preferred method because it allows collection of different information in a short period (Kumar & Ghodeswar, 2015) and suitable for field study. To test and quantify the relationships hypothesized, a questionnaire-based survey approach was adopted and a 27-items questionnaire covering six constructs was developed based on literature. The questionnaire was divided into two parts, the first part examined the various factors mentioned in the literature related to ECSR, and the items were adopted from different studies in the appropriate field and second section examined questions related to ECSR practices. Three items were used to measure environmental regulation (Williamson et al., 2006 based on grounded theory; Walls, Phan, & Berrone, 2011 based on Hart [1995] natural resource based view framework); three items are used to measure integration of CSR into corporate policy (Fowler & Hope, 2007 based on Hart [1995] natural resource-based view framework; Galbreath, 2009); two items are used to measure priority of board (Rangan, Chase, & Karim, 2012); three items were used to measure stakeholder involvement (Logistics Social Responsibility framework by Fineman & Clarke, 1996; Miao, Cai, & Xu, 2012); four items were used to measure responsible leadership (Scale of Discursive Responsible leadership by Hung, 2011; Voegtlin, 2011); and four items are used to measure partnering for CSR (Ağan, Kuzey, Acar, & Açıkgöz, 2016; Seitanidi & Crane, 2009). All measurements in the study were subjective assessment by respondents using 7-point Likert scales (where 1 stands for strongly disagree and 7 stands for strongly agree) the proposed framework for the study is presented in figure 1.

The scale was further validated with a panel of two academic and two corporate experts with an average of 12 years of experience in the area of CSR and provided comments on draft questionnaire, assess whether it is easy to understand and has distinctive meaning. The responses were further collected from the member of the CSR committee at manufacturing SMEs, and the demographic characteristics of these respondents like the type of ownership, industry and geographical location were also included.

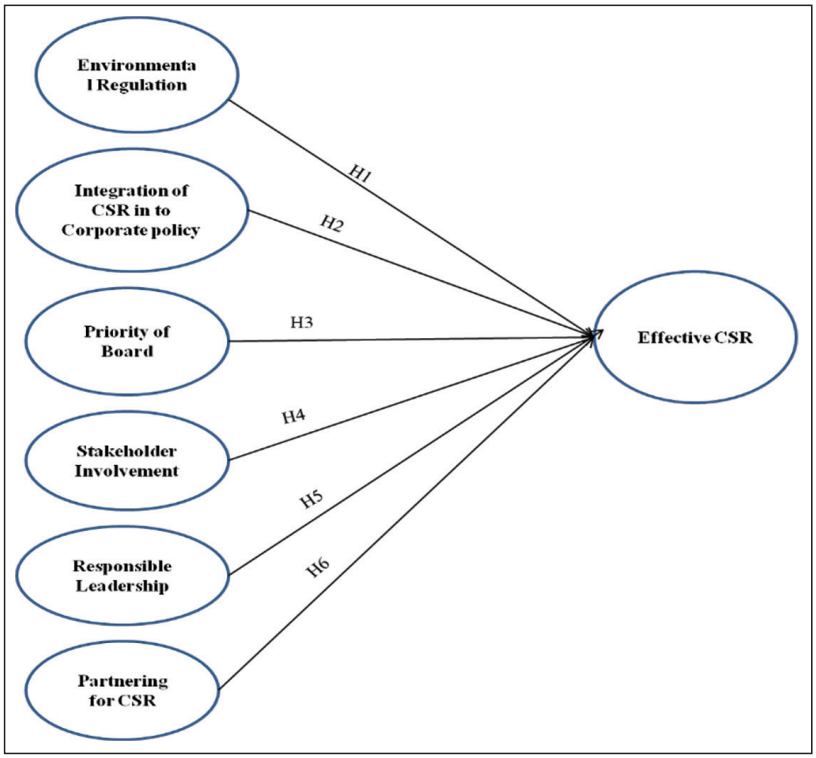


Figure 1. Proposed Framework

Source: The authors.

The study was conducted in the area of National Capital Region, India from February 2017 to April 2018. A total number of 361 manufacturing units are registered as SMEs at four areas under the National Capital Region, and all were considered for further examination. The respondents were the members of CSR committees who are generally middle- or high-level managers/directors in the organization (Table 2) and belonging to different industries (Figure 2). The respondents were contacted by email or by telephone and were asked to provide the relevant information through the questionnaire. The respondents through telephonic conversation were persuaded to participate in the study and were assured that their identity and information provided shall strictly be kept confidential and to be used for academic purpose only. For those who agreed to participate, a suitable time was decided with an appointment for data collection telephonically, or the questionnaire was sent to the email id in case there is no availability of suitable time slot. The response rate for the questionnaire was 44 per cent, that is, 158 responses were received from manufacturing SMEs, which is considered to be adequate as per Adlakha and Shukla (2014).

Table 2. Demographic Profile of Organizations

Respondent Level		Age of Organization (in years)	
Middle	39%	0–10	42%
Senior	51%	11–30	30%
Top management	10%	31–50	23%
		50 and above	4%

Source: The authors.

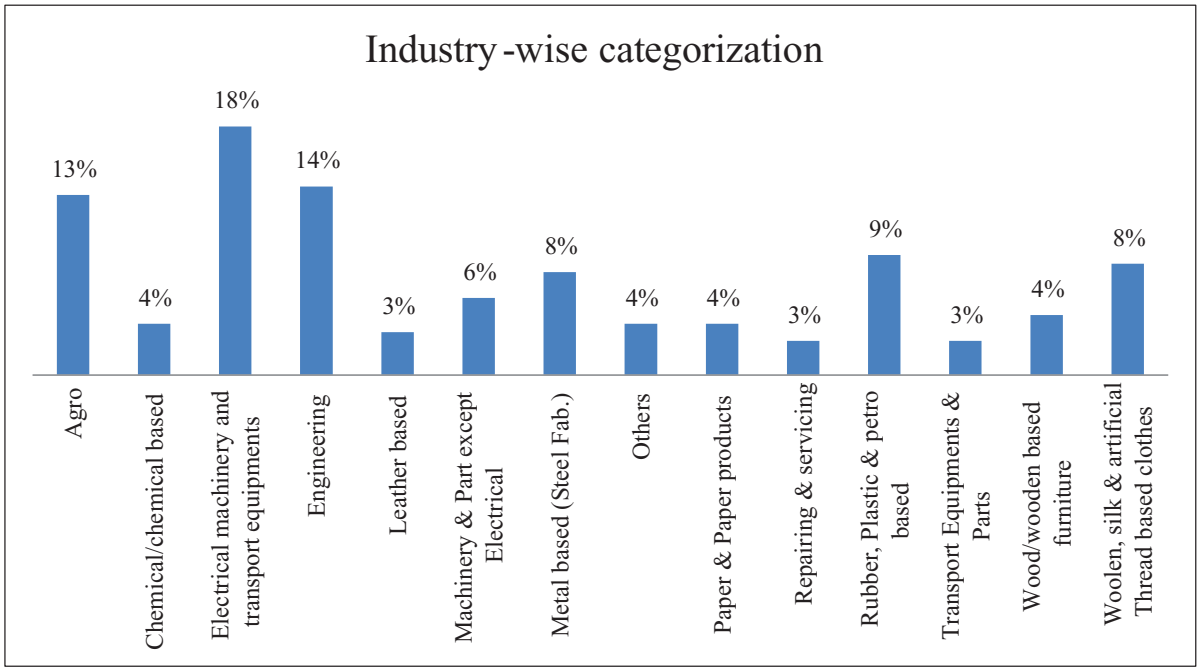


Figure 2.

Source: The authors.

The data provided were further analyzed using exploratory factor analysis to identify and validate the items that are contributing to every construct. A model fit was estimated using confirmatory factor analysis, and further hypotheses were tested using partial least square regression (PLSR) to explore the factors of ECSR. The tests were further interpreted based on the evidence from the literature.

Analysis and Findings

The data collected in the study from SMEs were tabulated in MS excel, and PLSR was used for analyzing the data. The research framework presented was analyzed using ADANCO 2.0.1. PLSR is a variance based method and the purpose of using PLSR in the study lies in the fact that no assumption regarding the distribution of data (Chin, Marcolin, & Newsted, 2003) and the method is also suitable for analysis based on small sample size (Ringle, Sarstedt, & Straub, 2012). Moreover, the indicators in the proposed model are reflective as they consider the effect of latent variables (Bollen & Lennox, 1991). The data collected were tabulated in MS excel before analyzing, and the reliability of the questionnaire was calculated followed by factor analysis and hypothesis testing. The findings of the study are elaborated subsequently. The validation of the PLS model is a two-step process:

1. Assessment of outer path model.
2. Assessing the inner path model.

The reliability and validity of the outer path model are to be established before the inner model is examined (Henseler, Ringle, & Sinkovics, 2009).

Reliability and Factor Analysis

The reliability analysis was measured using Cronbach's alpha (Cronbach, 1951). Further examination includes composite reliability (CR) (Werts et al., 1974), examination of item loading (Cramines & Zeller, 1979) and average variance extracted (AVE) (Fornell & Larcker, 1981). The scale was factor analyzed and items with the score greater than 0.7 were retained (Table 3). It is a common rule of thumb to indicate the convergent validity that item should load highly on their respective constructs than on other constructs (Yoo & Alavi, 2001).

Table 3. Factor Loadings and Cronbach's Alpha

Indicator	Items	Loadings	Construct	Cronbach's Alpha (α)
SI1	Educating customer programme	0.843	Stakeholder involvement	0.846
SI2	Employee awareness of CSR policy	0.882		
SI3	Environmental criteria for supplier selection	0.887		
CP1	Increase in transparency in organization	0.902	Integration of corporate policy	0.882
CP2	Incorporating ethics in policies	0.919		
CP3	Incorporating responsibility practices in organizational conduct	0.871		
RL1	Involving stakeholders in decision-making	0.933	Responsible leadership	0.967
RL2	Consider stakeholder claim	0.946		
RL3	Consensus among affected stakeholder	0.959		
RL4	Focus on social participation role	0.975		
PB1	Setting up special department for CSR	0.967	Priority of board	0.899
PB2	Training employees for CSR	0.937		
EG1	Environmentally friendly process and procedures	0.775	Environmental regulations	0.772
EG2	Environmental management system	0.807		
EG3	Code of conduct for self-governance	0.896		
PC1	Employee participation in voluntary activities	0.774	Partnering for CSR	0.717
PC2	Supporting NGO by charity	0.781		
PC3	Partnership with NGO	0.842		
CSR1	Responsible and ethical organization	0.796	ECSR	0.951
CSR2	Achieving SD through activities	0.834		
CSR3	Interaction of management with stakeholders	0.887		
CSR4	Self-performance of CSR activities	0.924		
CSR5	Providing financial support	0.799		

(Table 3 Continued)

(Table 3 Continued)

Indicator	Items	Loadings	Construct	Cronbach's Alpha (α)
CSR6	Inculcating CSR in activities of other department	0.817		
CSR7	Deciding activities by assessing local needs of society	0.896		
CSR8	Deciding activities by government 5-year plan	0.769		
CSR9	Practising organizational justice	0.891		

Source: The authors.

The items were grouped into seven constructs, and they all accounted for 83.85 per cent of the total variance. For testing the stability of scale, confirmatory factor analysis was employed on the information gathered through the questionnaire. The first step is to measure the construct validity, that is, how well the experiment measures up to its claim, and it is estimated through convergent validity and divergent validity. Convergent validity is the degree to which the measures of the constructs are related to each other. It can be measured using AVE that should exceed 0.5. All the values of AVE for the current framework varied from 0.63 to 0.91 (Table 4). The composite reliabilities varied from 0.62 to 0.97, which exceeded the recommended level of 0.6 (Aubert, Rivard, & Patry, 1996; Lawson-Body & Limayem, 2004).

Discriminant or divergent validity the measure in the construct is unrelated or distinct from each other. The result of discriminant validity is presented in Table 5 through the Fornell–Larcker criteria, that is, the square root of each construct's AVE should be higher than all the correlation with other constructs and the results obtained are satisfactory. Using values of fit indices a comparison for conceptual models and reject alternative models. Standardized root mean square residual (SRMR) is an absolute measure of fit, and it is the difference between observed and predicted correlation. The value less than 0.09 is considered to be a good fit (Hu & Bentler, 1999). The value of SRMR for the current framework obtained was 0.08, and it is a satisfactory level of fit. Thus, it can be inferred that ECSR is best identified along the six dimensions.

Table 4. Convergent Validity

Construct	Average Variance Extracted (AVE)	Composite Reliability (CR)
SI	0.758	0.904
RL	0.909	0.976
EG	0.684	0.759
CSR	0.718	0.958
PB	0.906	0.624
CP	0.805	0.925
PC	0.640	0.734

Source: The authors.

The variable identified from literature holds support in the empirical investigation. The first factor *Stakeholder involvement* explains the perception of the organization towards involving stakeholder in the CSR practices to make them effective. It also elaborates on the engaging different stakeholders (like employees, suppliers and customers) in responsible practices that are rewarded mutually. The second factor *Integration of CSR into Corporate policy* is related to developing a corporate code of conduct on ethics and incorporating the ethical codes into mission, vision, strategic issues and practices to develop the competitive advantage for the organization. The third factor *Environmental Regulations* elaborates on concerns for environmental sustainability as a driver for ECSR. Environmental protection and conservation are the important aspects of CSR policies of corporations. The fourth factor *Priority of Board towards CSR* is related to keeping track of wide-ranging social, environmental and ethical issues to ensure right CSR priorities for the organization. The fifth factor *Responsible Leadership* is related to role and responsibilities of business leaders that centres attention on responsibility, accountability and moral decision-making for those who are affected and have stakes in the business. The sixth factor *Partnering for CSR* describes alliances organization are attending to address the societal problem through non-profit partnerships.

Hypothesis Testing

To test the proposed hypothesis, the measurement model was converted into the structural model in ADANCO. The results are discussed and interpreted based on regression weight and Cohen's f^2 (Table 6), and they indicate that the effect size on the latent variable and the value greater than $f^2 \geq 0.35$ has a strong effect (Cohen, 1992). The first relationship between supporting environmental regulations and ECSR is found statistically significant ($p < 0.05$). This supports *H1* that environmental regulations positively predisposition towards ECSR and based on f^2 it can be inferred that it has the strongest direct effect on the effectiveness of CSR in small and medium manufacturing companies. In other words, it indicates that manufacturing organizations have a stronger inclination towards undertaking environmentally friendly practices as a part of CSR policy and it also positively influence its performance. The finding is consistent with Cetindamar and Husoy (2007) and Sethi, Martell, and Demir (2017), as they tested that there is a positive impact of environmental responsible CSR practices on the long-run and short-run performance of the firm.

Table 5. Discriminant Validity by Fornell–Larcker Criterion

Construct	SI	RL	EG	CSR	PB	CP	PC
SI	0.758						
RL	0.650	0.909					
EG	0.247	0.308	0.684				
CSR	0.060	0.215	0.652	0.718			
PB	0.680	0.672	0.110	0.010	0.906		
CP	0.500	0.782	0.298	0.228	0.526	0.805	
PC	0.081	0.136	0.483	0.628	0.008	0.129	0.639

Source: The authors.

Note: Squared correlations; AVE in the diagonal.

Table 6. Results of Hypothesis Testing

Effect	Original Coefficient	Mean Value	Standard Error	*t*-Value	*p*-Value (2-sided)	Cohen's f^2
SI → ECSR	−0.342	−0.336	0.082	−4.194	0.000	0.179
RL → ECSR	0.339	0.327	0.096	3.529	0.000	0.098
EG → ECSR	0.502	0.501	0.046	10.967	0.000	0.647
PB → ECSR	−0.259	−0.243	0.115	−2.246	0.025	0.086
CP → ECSR	0.227	0.224	0.067	3.3890	0.0001	0.067
PC → ECSR	0.358	0.364	0.0489	7.349	0.0000	0.381

Source: The authors.

The relationship between the integration of CSR into corporate policy and ECSR is found to be statistically significant ($p < 0.05$). This supports *H2*, indicating that making CSR a part of compliance structures and processes is critical in ensuring the commitments. The finding is consistent to Morimoto, Ash, and Hope (2005), Castka, Balzarova, Bamber, and Sharp (2004) and Laguir, Laguir, and Tchemeni (2019); as they found that integration of CSR into corporate policy is one of the critical achievements of successful CSR.

The relationship between the priority of Board of Directors and ECSR is found to be statistically significant ($p < 0.05$), and the value of f^2 examines its impact as moderate. It supports *H3*; indicating board preparing guidelines for implementing the priority activity is associated to nature of business cannot positively integrate with effective social practices in manufacturing companies. The finding is consistent with Ibrahim, Howard, and Angelidis (2003), Arena, Bozzolan, and Michelon (2015) and García-Sánchez, Gómez-Miranda, David, and Rodríguez-Ariza (2019); as they founded that the directors have the weaker orientation towards discretionary responsibilities.

The relationship between stakeholder involvement and ECSR is found to be statistically significant ($p < 0.05$). It supports *H4*, indicating higher exposure of the company to social responsibilities more likely they are to attract critical attention to the stakeholder. The finding is consistent with Ashforth and Gibb (1990), Szczanowicz and Saniuk (2014) and Margolis and Walsh (2003), as they explored the previous research and highlighted two opposite views socio-economic world that cannot be positively integrated.

The relationship between responsible leadership and ECSR is found to be statistically significant ($p < 0.05$). It supports *H5*, indicating that leading by values is a core management philosophy as a practice of corporate responsibility. The finding is consistent with the Székely and Knirsch (2005) and Coulson-Thomas (2018); as they used their metrics at various companies and found a strong association between the responsible style of leadership and social responsibility practices.

The relationship between partnering for CSR and ECSR is found to be statistically significant ($p < 0.05$). It supports *H6*, indicating that implementing alliances is a suitable platform for discharging social responsibility. The finding is consistent with Albareda, Lozano, Tencati, Midttun, and Perrini (2008), DeMasi (2011) and Seitanidi and Crane (2009); as they found partnership for CSR as a driver for a successful CSR programme.

Thus, ECSR was significantly determined by environmental regulations (EG), incorporation of CSR into to corporate policy (CP), CSR as a priority of board of directors (PB), stakeholder holder involvement (SI), responsible leadership (RL) and partnering for CSR (PC), resulting in $R^2 = 0.845$. In other words, the above-mentioned variables explain 84.5 per cent of the variance of ECSR.

Discussion and Implication

CSR practices including monetary donations, implementing environmentally friendly practices at workplace describes the companies' efforts to improve society. The CSR mandate in 2014 and increase in the sustainability consciousness of stakeholders has made companies design and undertake CSR programme seriously as any false claim related to social change could lead to negative effects on its performance. So, with the gradual progress in research in the domain of CSR and sustainability practices of organizations, recent studies have focussed on how to achieve effectiveness in social responsibility practices. In this direction, this study has attempted to explore the factors that lead to effective implementation of social responsibility policy through Gandhi's trusteeship theory. The findings of the study indicate that in the case of the manufacturing SMEs the strongest drivers for ECSR practices are environmental regulations and partnering for CSR practices. These companies are strongly concerned with the environmentally friendly practices, environmental management systems and environmental code of self-conduct as they realize their responsibility towards the environment. They also believe in discharging their responsibilities through productive partnerships by involving with NGO's activities at ground level, or by just providing monetary support to NGO's activities and also encouraging the employees to involve themselves in social activities on the voluntary basis. So, it can be inferred that the success of CSR policy largely depends on the following environmental regulations adhered and getting into the partnership for discharging CSR activities.

The current study tested the impact of several variables on ECSR that has a significant theoretical contribution. The first considerable relationship of supporting environmental regulations and ECSR confirms the organizations' decision, influenced by free market value, to undertake related environmental practices. It indicates implementation of environment-friendly practices within the organization, consuming resources responsibly and deliberate continuous and conscious evaluation of environment. It also provides a suitable yardstick important to measure if companies' pursuing environmental activities are a part of compliance or something beyond it. Our findings imply CSR policy as the environment is the key component for SD. The research related to the impact of environmental regulations on ECSR in manufacturing SMEs is consistent with the results of Williamson et al. (2006), Dean et al. (2000) and OECD (Gordon, 2001). Companies with such environment-oriented approach seem meaningful, and this drives their environmental responsibility.

Further, the significant relationship between integration of CSR into corporate policy and ECSR confirms that manufacturing SMEs are incorporating the essence of responsibility into overall organizational systems this highlights strategic intent towards CSR. SMEs are implementing the critical task by increasing transparency through policy sharing with stakeholders (e.g., employees and suppliers), developing organizational ethical principles and incorporating them in policies and conducting affairs responsibly to create sustainable value. Although, as per our analysis integration of CSR into corporate policy has a weak effect on the effectiveness of CSR. Thus, this result can be explained in other words as the fact that incorporation of CSR approach in every aspect of the business may enable the organization to gain competitive advantage but leaders of these companies as the obstacle to gain notoriety by the enterprises. Indeed, the incorporation of CSR approach may have some economic benefit appropriate for implementing SMEs. The findings of this study have implications to CSR practices as it can be better defined and their impact on the overall policy can be assessed. Our result of the investigation is consistent with Boubakary and Moskplai (2016). SMEs implementing CSR with such approach can effectively contribute to sustainability from all its activities.

Also, the relationship between Stakeholder Involvement and ECSR confirms that Stakeholders have a negative perception of CSR activities at SMEs. Stakeholders include a wide range of parties, such as

employees, customers, suppliers, NGO's and local communities, because each of them shares a different relationship, and hence their engagement shall depend upon the structures of the organization. In the current study, SMEs are attempting to engage stakeholders by undertaking a customer education programme, communicating with employees about organizations' CSR policy and developing criteria for supplier selection by the environmental assessment. The analysis highlights that stakeholder involvement and ECSR are opposite to each other, as stakeholder engagement in SMEs can create more work and complicate their exiting process which manufacturing companies of small and medium scale may not be able to support out of their existing capital. Thus, from stakeholder perspective business is about adding value to all stakeholders and create good for them (Freeman & Gilbert, 1992; Freeman & Velamuri, 2006) and they are opposite to shareholder expectations (Margolis & Walsh, 2003). The findings of the study have implications for CSR practitioners for deciding on the activities to be undertaken from the stakeholder perspective in manufacturing companies. The results of the study are consistent with Parmar et al. (2010) that stakeholders are not enthusiastic about CSR as it said to reproduce separation thinking.

Another relationship was studied between responsible leadership and ECSR. The significant association suggests that governance and leadership style is essential for the organization to discharge their social responsibility effectively as involve, recognize and motivate the employees in the firms. SMEs engaged in manufacturing are implementing the agenda of responsible leadership through top management considering stakeholders claim and involving them in decision-making, achieving consensus among them and broadly working towards social participation. So, managers of SMEs can refer to the findings of the study to implement the concept of responsible leadership. The results of the study are consistent with that of Groves and LaRocca (2011). The managers of SMEs today are paying close attention to ethically oriented investors, and sustainability-focussed think tanks to make their worth in community.

Further, the relationship between partnering for CSR and ECSR is statistically significant, and it is one key of the factor for SMEs in discharging their responsibility effectively as it can help them in tackling social challenges through the expertise of public, private and non-profit partners. Finding a right partner is critical for the success of organizations' CSR strategy as it is expected to co-create value for both the organization and society. SMEs are implementing social responsibility partnership by encouraging employees to take up voluntary activities, giving donations to NGOs and creating associations with the non-profit partner for societal activities. The findings of the study have implications for CSR practitioners at SMEs while framing their CSR policy. The results of the study are consistent with Seitanidi and Crane (2009). The CSR partnership results from outcome-centric approach through effective monitoring of social projects and removal of malpractices. Thus, playing off with each other's strength can foster a shared sense of responsibility that can effectively increase the social value of projects.

Thus, this study contributes towards the understanding of factors that leads to ECSR in Indian manufacturing SMEs. This study elaborates on the relevance of CSR activities in the SME context and their perspective towards making CSR policy effective. Compared to previous studies, this study prominently addresses the factors that contribute to the effectiveness of social responsibility in manufacturing SMEs. In other words, environmental regulations and partnering for CSR are essential factors from SMEs perspective that make CSR activities effective. Further, the study recognizes the role of responsible leadership and stakeholder involvement in making CSR decisions impactful. The result is in line with the findings from the literature that explains the purpose of environmental regulations, integration of CSR into corporate policy, priority of board, stakeholder involvement, responsible

leadership and partnering for CSR in determining ECSR. The current study is different from previous studies on ECSR in explaining the factors that make the social responsibility of practices of SMEs more impactful for the society at ground level rather than just focussing on organizational performance.

Conclusion

CSR activities are expected to encourage voluntary culture and managerial developments that will create the basis for SD. The current study highlights that social responsibility and environmental regulations are becoming an essential part of regular practice at SMEs. The findings from our research can be summarized as Indian SMEs are actively adopting focussed environmental CSR as it is the valued driver with many benefits that are not reflected in financial terms. Since these benefits have high social meaning SMEs are more likely to have favourable inclination towards environmental CSR.

Second, supporting responsible leadership, stakeholder involvement, integration of CSR into corporate policy and priority of board towards CSR enhances the ability of organizations to address social issues and achieve competitive advantage. It reflects that SMEs are concerned with the interest of society along with its shareholders. Hence, the responsible strategy not only fulfils the SD matters but also addresses economic concerns, therefore, enhancing the effectiveness of CSR. Even though the current study has highlighted a conceit in the relationships as mentioned earlier as few SMEs under investigation have CSR orientation since the beginning, others implemented an accelerated CSR in corporate strategy.

Third, non-profit partnerships have become an alternate strategy for CSR. Its recent proliferation in recent years is due to the potential benefit that arises from synergies that can ensue. The current study has shown the positive connection between partnering for CSR and its effectiveness as the partnership between SMEs and NGOs are formed in pursuit of CSR. These partnerships are taking different forms by employee volunteering, monetary support and joint ventures in light of rising expectations.

Fourth, the study has implications for CSR professionals in developing CSR strategies for SMEs in the Indian context. It is evident from our findings that trusteeship theory by Mahatma Gandhi is relevant to corporate responsibility as the approach works on self-interest along with pro-social behaviour. The results of the study have also outlined the contribution of trusteeship theory as a framework for ECSR.

Thus, the study enhances the understanding of ECSR in the Indian context and also offers insight to understand the SMEs' viewpoint on CSR in the Indian market. The findings can be used by managers heading CSR department who are interested to know the multifariousness of CSR activities. Thus, professionals can effectively use them so that SMEs can address social issues and at the same time maintain their market share.

Limitations

The current work is a starting point for empirical studies, and it has certain limitations. The sample data is collected from SMEs in the National Capital Region and may not represent the entire SMEs in India *per se*. Generalization of the results to the different country under contrasting institutional settings should be made with caution. Thus, the current work has sought analytical generalization rather than statistical generalization.

Future Research

The current study has paved the way for further research. First, the basis of ECSR in current study only captures trusteeship theory. Further investigations can be made based on other theoretical frameworks of CSR to explore the CSR initiatives that 'pay off'. Second, the effectiveness of CSR can also be examined between CSR practices of domestic SMEs and SMEs with international associations. Third, it would also be interesting to consider SMEs communication about CSR in the existing model of CSR implementation to explore its impact on the effectiveness of social responsibility practices. Finally, a cross-country comparison can be pursued to address the issue of generalizability resulting from focus on one country.

Acknowledgement

The authors are indebted to anonymous referees for their extremely useful suggestions for improving the quality of article. Usual disclaimers apply.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

References

Adams, C. (2013). The role of leadership and governance in transformational change towards sustainability. *Global Responsibility ,* (9), 9–12.

Adeneye, Y. B., & Ahmed, M. (2015). Corporate social responsibility and company performance. *Journal of Business Studies Quarterly, 7*(1), 151–166.

Adlakha, A., & Shukla, V. (2014). Impact of Organized Retail Outlets on Unorganized Small Shops and Hawkers in Uttarakhand. *Asian Journal of Research in Business Economics and Management, 4*(3), 171.

Ağan, Y., Kuzey, C., Acar, M. F., & Açıkgöz, A. (2016). The relationships between corporate social responsibility, environmental supplier development, and firm performance. *Journal of Cleaner Production, 112*, 1872–1881.

Albareda, L., Lozano, J. M., Tencati, A., Midttun, A., & Perrini, F. (2008). The changing role of governments in corporate social responsibility: Drivers and responses. *Business Ethics: A European Review, 17*(4), 347–363.

Alvarado-Herrera, A., Bigne, E., Aldas-Manzano, J., & Curras-Perez, R. (2017). A scale for measuring consumer perceptions of corporate social responsibility following the sustainable development paradigm. *Journal of Business Ethics, 140*(2), 243–262.

Álvarez-González, L. I., García-Rodríguez, N., Rey-García, M., & Sanzo-Perez, M. J. (2017). Business-nonprofit partnerships as a driver of internal marketing in nonprofit organizations. Consequences for nonprofit performance and moderators. *BRQ Business Research Quarterly, 20*(2), 112–123.

Arena, C., Bozzolan, S., & Michelon, G. (2015). Environmental reporting: Transparency to stakeholders or stakeholder manipulation? An analysis of disclosure tone and the role of the board of directors. *Corporate Social Responsibility and Environmental Management, 22*(6), 346–361.

Arya, B., & Salk, J. E. (2006). Cross-sector alliance learning and effectiveness of voluntary codes of corporate social responsibility. *Business Ethics Quarterly, 16*(2), 211–234.

Ashforth, B. E., & Gibbs, B. W. (1990). The double edge of organizational legitimation. *Organization Science, 1*(2), 177–194.

Aubert, B. A., Rivard, S., & Patry, M. (1996). Development of measures to assess dimensions of IS operation transactions. *Omega, International Journal of Management Science, 24*(6), 661–680.

Axelrod, A. (2010). Gandhi, CEO: 14 *Principles to Guide & Inspire Modern Leaders.* Sterling.

Babatunde, J. H., & Adeyemi, A. A. (2015). The effectiveness of Islamic banks' corporate social responsibility as perceived by gender in Malaysia: An exploratory research. *Journal of Creative Writing, 1*(2), 1–18.

Bader, G. (1995). Gandhi on trusteeship: A transforming ethic. *World Business Academy Perspectives, 9*(4), 41–50.

Bajaj, R. (2016). From trusteeship model to the new vehicles of CSR. In *The privilege of responsibility: Perspectives from corporate India.* CII. Retrieved from https://www.ciiblog.in/industry/from-the-trusteeship-model-to-the-new-vehicles-of-csr/

Balakrishnan, J., Malhotra, A., & Falkenberg, L. (2017). Multi-level corporate responsibility: A comparison of Gandhi's trusteeship with stakeholder and stewardship frameworks. *Journal of Business Ethics, 141*(1), 133–150.

Balasubramanian, N. (2010). Governing the socially responsible corporation: A Gandhian perspective. In A. Gupta (Ed.), *Ethics, business and society: Managing responsibly* (pp. 157–181). New Delhi, India: SAGE Publications.

Barross-Mendez, M. J., Galera-Casquet, C., & Valero-Amaro, V. (2015). Proposal of a social alliance success model from a relationship marketing perspective: A meta-analytical study of the theoretical foundations. *Business Research Quarterly, 18*(3), 188–203.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of management, 17*(1), 99–120.

Baumann-Pauly, D., Wickert, C., Spence, L. J., & Scherer, A. G. (2013). Organizing corporate social responsibility in small and large firms: Size matters. *Journal of Business Ethics, 115*(4), 693–705.

Batool, S., Butt, A., & Niazi, B. (2016). Types of Corporate Social Responsibility Practices across the Industry in Pakistan and Their Effectiveness. In *Corporate Responsibility and Stakeholding* (pp. 225-246). Emerald Group Publishing Limited.

Beck, T., Demirguc-Kunt, A., & Martinez Peria, M. S. (2008). *Bank financing for SMEs around the world: Drivers, obstacles, business models, and lending practices*. Washington, DC: The World Bank.

Bimal, A., & Puranik, R. (2004). A review of corporate social responsibility in India. *Development, 47*(3), 93–100.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*(2), 305.

Boubakary, D., & Moskolaï, D. D. (2016). The Influence of the Implementation of CSR on Business Strategy: An Empirical Approach based on Cameroonian Enterprise. *Arab Economic and Business Journal, 11*(2), 162–171.

Brammer, S., & Millington, A. (2003). The effect of stakeholder preferences, organizational structure and industry type on corporate community involvement. *Journal of Business Ethics, 45*(3), 213–226.

Brammer, S., & Pavelin, S. (2004). Building a good reputation. *European Management Journal, 22*(6), 704–713.

Brown, M. E., & Treviño, L. K. (2006). Ethical leadership: A review and future directions. *The Leadership Quarterly, 17*(6), 595–616.

Caldwell, C., Bischoff, S. J., & Karri, R. (2002). The four umpires: A paradigm for ethical leadership. *Journal of Business Ethics, 36*(1–2), 153–163.

Cameron, K. (2011). Responsible leadership as virtuous leadership. In *Responsible Leadership* (pp. 25–35). Dordrecht: Springer.

Caldwell, C., Hayes, L. A., Bernal, P., & Karri, R. (2008). Ethical stewardship–implications for leadership and trust. *Journal of Business Ethics, 78*(1–2), 153–164.

Cannon, T. (1994). *Corporate responsibility.* London, UK: Pitman.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment* (Vol. 17). SAGE publications.

Carroll, A. B. (1979). A Three-Dimensional Model of Corporate Performance. *Academy of Management Review, 4*(4), 497–505.

_____. (1983). Corporate social responsibility: Will industry respond to cut-backs in social program funding? Vital Speeches of the Day, 49, 604–608.

_____. (1991). The pyramid of corporate social responsibility: Toward the moral management of organizational stakeholders. *Business Horizons, 34*(4), 39–48.

Carroll, A. B., & Buchholtz, A. K. (2012). *Business & society: Ethics, sustainability and stakeholder management* (8th ed.). Mason, OH: South-Western Cengage Learning.

Carroll, A. B., Lipartito, K. J., Post, J. E., & Werhane, P. H. (2012). *Corporate responsibility: The American experience*. Cambridge, UK: Cambridge University Press.

Castka, P. (2004). *Implementing effective corporate social responsibility and corporate governance: A framework*. BSI British Standards Institution.

Castka, P., Balzarova, M. A., Bamber, C. J., & Sharp, J. M. (2004). How can SMEs effectively implement the CSR agenda? A UK case study perspective. *Corporate Social Responsibility and Environmental Management, 11*(3), 140–149.

Cetindamar, D., & Husoy, K. (2007). Corporate social responsibility practices and environmentally responsible behavior: The case of the United Nations global compact. *Journal of Business Ethics, 76*(2), 163–176.

Chebet, R. G., & Muturi, W. (2018). Effect of corporate social responsibility on organizational effect of corporate social responsibility on organizational. *International Journal of Social Sciences and Information Technology, IV*(II), 50–62.

Chin, W. W., Marcolin, B. L., & Newsted, P. R. (2003). A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study. *Information Systems Research, 14*(2), 189–217.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155.

Collier, J., & Estebann, R. (2007). Corporate social responsibility and employee commitment. *Business Ethics: A European Review, 16*(1), 19–33.

Coulson-Thomas, C. (2018). Corporate leadership and governance for increasing stakeholder involvement and developing stronger connections. *Effective Executive, 21*(1), 7–25.

Crane, A., Matten, D., & Spence, L. J. (2014). *Corporate social responsibility: Readings and cases in a global context*. London, UK: Routledge.

Crespin, R. (2012, February 10). *Where CSR fits on the board's agenda*. Retrieved from https://www.forbes.com/sites/csr/2012/02/10/where-csr-fits-on-the-boards-agenda/#3a46fb4879c8

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Daptardar, V. (2012). Gandhian relevance to environmental sustainability. In *Gandhi in the new millennium—Issues and challenges*. Mumbai: Khandwala Publishing House.

Darcy, C., Hill, J., McCabe, T. J., & McGovern, P. (2014). A consideration of organisational sustainability in the SME context: A resource-based view and composite model. *European Journal of Training and Development, 38*(5), 398–414.

Dawar, G., & Singh, S. (2016). Corporate social responsibility and gender diversity: A literature review. *Journal of IMS Group, 13*(1), 61–71.

De Clercq, D., & Voronov, M. (2011). Sustainability in entrepreneurship: A tale of two logics. *International Small Business Journal, 29*(4), 322–344.

Dean, T. J., Brown, R. L., & Stango, V. (2000). Environmental regulation as a barrier to the formation of small manufacturing establishments: A longitudinal examination. J*ournal of Environmental Economics and Management, 40*(1), 56–75.

DeMasi, E. (2011). A comparative analysis of corporate social responsibility in commercial banks: Case studies from the United States and the United Kingdom (Master's thesis, Duquesne University). Retrieved from https://dsc.duq.edu/etd/477.

Chrun, E., Dolšak, N., & Prakash, A. (2016). Corporate environmentalism: Motivations and mechanisms. *Annual Review of Environment and Resources, 41*, 341–362.

Donaldson, L., & Davis, J. H. (1991). Stewardship theory or agency theory: CEO governance and shareholder returns. *Australian Journal of Management, 16*(1), 49–64.

Donaldson, T., & Preston, L. (1995). The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of Management Review, 20*(1), 65–91.

Eilbirt, H., & Parket, I. R. (1973). The practice of business: The current status of corporate social responsibility. *Business horizons, 16*(4), 5–14.

Elankumaran, S., Seal, R., & Hashmi, A. (2005). Transcending transformation: Enlightening endeavours at Tata Steel. *Journal of Business Ethics, 59*(1–2), 109–119.

Ethics in Action Awards. (2003). What is Corporate Social Responsibility?

Feil, M. (2012). *Global governance and corporate social responsibility in conflict zones*. Basingstoke, UK: Palgrave Macmillan.

Fineman, S., & Clarke, K. (1996). Green stakeholders: Industry interpretations and response. *Journal of Management studies, 33*(6), 715–730.

Florida, R., & Davison, D. (2001). Gaining from green management: environmental management systems inside and outside the factory. *California management review, 43*(3), 64–84.

Fowler, S. J., & Hope, C. (2007). Incorporating sustainable business practices into company strategy. *Business Strategy and the Environment, 16*(1), 26–38.

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics.

Freeman, R. E. (1984). *Strategic Management: A Stakeholder Approach*. Boston: Pitman.

Freeman, R. E., & Gilbert, D. R. (1992). Business, ethics, and society: A critical agenda. *Business and Society, 31*(1), 9–17.

Freeman, R. E., Harrison, J. S., Wicks, A. C., Parmar, B. L., & De Colle, S. (2010). *Stakeholder theory: The state of the art*. Cambridge University Press.

Freeman, R. E., & McVea, J. (2001). A stakeholder approach to strategic management. Retrieved from http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=263511

Freeman, R. E., & Velamuri, S. R. (2006). A new approach to CSR: Company stakeholder responsibility. In M. Morsing & A. Kakabadse (Eds.), *Corporate social responsibility: From aspiration to application* (pp. 9–23). Basingstoke, UK and Hampshire: Palgrave Macmillan.

_____(2008). A New Approach to CSR: Company Stakeholder Responsibility, Prieiga. Retrieved from http://ssrn.com/abstract, 1186223

Frederick, W.C. (1960). The growing concern over business responsibility. *California Management Review, 2*(4), 54–61.

Friedman, M. (1970). A Friedman doctrine: The social responsibility of business is to increase its profits. *The New York Times Magazine, 13*(1970), 32–33.

Galbreath, J. (2009). Building corporate social responsibility into strategy. *European Business Review, 21*(2), 109–127.

García-Sánchez, I. M., Gómez-Miranda, M. E., David, F., & Rodríguez-Ariza, L. (2019). Board independence and GRI-IFC performance standards: The mediating effect of the CSR committee. *Journal of Cleaner Production, 225*, 554–562.

Gautam, R., & Singh, A. (2010). Corporate social responsibility practices in India: A study of top 500 companies. *Global Business and Management Research: An International Journal, 2*(1), 41–56.

Ghosh, S. (1989). Trusteeship in industry: Gandhiji's dream and contemporary reality. *Indian Journal of Industrial Relations, 25*(1), 35–44.

Goodpaster, K. E. (2013). Tenacity: The American pursuit of corporate responsibility. *Business and Society Review, 118*(4), 577–605.

Gopinath, C. (2005). Trusteeship as a moral foundation for business. *Business and Society Review, 110*(3), 331–344.

Gordon, K. (2001). The OECD guidelines and other corporate responsibility instruments: A comparison (OECD Working Papers on International Investment No. 2001/05). OECD Publishing. Retrieved from http://dx.doi.org/10.1787/302255465771

Greenstone. (2014, December). *Why engaging with your stakeholders is important for CSR reporting*. Retrieved from https://www.greenstoneplus.com/blog/why-engaging-with-your-stakeholders-is-important-for-csr-reporting

Groves, K. S., & LaRocca, M. A. (2011). Responsible leadership outcomes via stakeholder CSR values: Testing a values-centered model of transformational leadership. *Journal of Business Ethics, 98*(Suppl 1), 37–55.

Hansen, E. G. (2008). Responsible leadership requires responsible leadership systems: The case of Merck Ltd., Thailand. Retrieved from https://ssrn.com/abstract=1440290 or http://dx.doi.org/10.2139/ssrn.1440290

Harris, J. D., Sapienza, H. J., & Bowie, N. E. (2009). Ethics and entrepreneurship. *Journal of Business Venturing, 24*(5), 407–418.

Hart, S. L. (1995). A natural-resource-based view of the firm. *Academy of management review, 20*(4), 986–1014.

Heismann, K. (2014, October). *Why stakeholder engagement is key to successful CSR programs*. Retrieved from https://www.greenbiz.com/blog/2014/10/23/stakeholder-engagement-key-csr-online-communities

Henseler, J., Ringle, C., & Sinkovics, R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing, 20*, 277–320.

Hernandez, M. (2008). Promoting stewardship behavior in organizations: A leadership model. *Journal of Business Ethics, 80*(1), 121–128.

Hopkins, M 1998. *The Planetary Bargain: Corporate Social Responsibility Comes of Age*. London: Macmillan.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Hung, H. (2011). Directors' roles in corporate social responsibility: A stakeholder perspective. *Journal of Business Ethics, 103*(3), 385–402.

Ibrahim, N. A., Howard, D. P., & Angelidis, J. P. (2003). Board members in the service industry: An empirical examination of the relationship between corporate social responsibility orientation and directorial type. *Journal of Business Ethics, 47*(4), 393–401.

Idemudia, U. (2009). Oil extraction and poverty reduction in the Niger delta: A critical examination of partnership initiatives. *Journal of Business Ethics, 90*(1), 91–116.

IIRC. (2013). The International IR Framework.

Jain, P., Vyas, V., & Chalasani, D. P. S. (2016). Corporate social responsibility and financial performance in SMEs: A structural equation modelling approach. *Global Business Review, 17*(3), 630–653.

Jaiyeoba, H. B., Adewale, A. A., & Quadry, M. O. (2018). Are Malaysian Islamic banks' corporate social responsibilities effective? A stakeholders' view. *International Journal of Bank Marketing, 36*(1), 111–125.

Jamali, D., Lund-Thomsen, P., & Jeppesen, S. (2017). SMEs and CSR in developing countries. *Business & Society, 56*(1), 11–22.

Jamali, D., Zanhour, M., & Keshishian, T. (2009). Peculiar strengths and relational attributes of SMEs in the context of CSR. *Journal of Business Ethics, 87*(3), 355–367.

Jenkins, H. (2004). A critique of conventional CSR theory: An SME perspective. *Journal of General Management, 29*(4), 37–57.

Joseph, S. K., Mahodaya, B., & Pradhan, R. (2016). *Trusteeship: A path less travelled*. Institute of Gandhian Studies.

Kansal, M., Joshi, M., & Batra, G. S. (2014). Determinants of corporate social responsibility disclosures: Evidence from India. *Advances in Accounting, 30*(1), 217–229.

Kapoor, A. (2005). Gandhi's trusteeship. In M. M. Sankhdher & G. Kaur (Eds.), *Politics in India: Ancient India, politics of change, modern India* (pp. 252–268). Delhi: Deep & Deep Publications.

Khan, S. (2008, August). Conceptualizing corporate social responsibility: An Indian perspective. In *Academy of Management Proceedings* (Vol. 2008, No. 1, pp. 1–6). Briarcliff Manor: Academy of Management.

Kim, N., Cho, E., Kim, Y., & Lee, M. (2011). Developing an effective strategic mix of corporate philanthropy. *The Service Industries Journal, 31*(7), 1049–1062.

Knowledge @Wharton. (2012, May). *How companies integrate CSR initiatives into everyday business*. Retrieved from http://parisinnovationreview.com/articles-en/how-companies-integrate-csr-initiatives-into-everyday-business

Kolge, N., & Sreekumar, N. (2011). Gandhi's criticism of industrialization and modernity: An environmental perspective. In S. K. Joseph & B. Mahodaya (Eds.), *Gandhi, environment and sustainable future*. Wardha, India: Institute of Gandhian Studies. Retrieved from http://www.gvpwardha.iecit.in/documents/books/env/4.pdf

Krisnawati, A., Yudoko, G., & Bangun, Y. R. (2018). Modeling an effective corporate social responsibility based on systems theory and management functions: A case study in Indonesia. *International Journal of Business and Society, 19*(2), 249–261.

Kumar, P., & Polonsky, M. J. (2017). An analysis of the green consumer domain within sustainability research: 1975 to 2014. *Australasian Marketing Journal (AMJ), 25*(2), 85–96.

Kumar, R., Murphy, D. F., & Balsari, V. (2001). *Altered images: The 2001 state of corporate responsibility in India poll. Understanding and encouraging corporate responsibility in South Asia: Update one*. Tata Energy Research Institute, Delhi.

Laguir, L., Laguir, I., & Tchemeni, E. (2019). Implementing CSR activities through management control systems: A formal and informal control perspective. *Accounting, Auditing & Accountability Journal, 32*(2), 531–555.

Lawson-Body, A., & Limayem, M. (2004). The impact of customer relationship management on customer loyalty: The moderating role of web site characteristics. *Journal of Computer-Mediated Communication, 9*(4).

López-Pérez, M. E., Melero, I., & Javier Sese, F. (2017). Management for sustainable development and its impact on firm value in the SME context: Does size matter? *Business Strategy and the Environment, 26*(7), 985–999.

Luetkenhorst, W. (2004). Corporate social responsibility and the development agenda. *Intereconomics, 39*(3), 157–166.

Lydenberg, S. D., Rogers, J., & Wood, D. (2010). *From transparency to performance: Industry-based sustainability reporting on key issues*. Cambridge, MA: Hauser Center for Nonprofit Organizations.

Maak, T. (2007). Responsible leadership, stakeholder engagement and the emergence of social capital. *Journal of Business Ethics, 74*(4), 329–343.

Maignan, I., & Ferrell, O. (2000). Measuring corporate citizenship in two countries: The case of the US and France. *Journal of Business Ethics, 23*(3), 283–297.

Mannell, J. (2010). Are the sectors compatible? International development work and lessons for a business-nonprofit partnership framework? *Journal of Applied Social Psychology, 40*(5), 1106–1122.

Margolis, J. D., & Walsh, J. P. (2003). Misery loves companies: Rethinking social initiatives by business. *Administrative Science Quarterly, 48*(2), 268–305.

McWilliams, A., Siegel, D. S., & Wright, P. M. (2006). Corporate social responsibility: Strategic implications. *Journal of Management Studies, 43*(1), 1–18.

Miao, Z., Cai, S., & Xu, D. (2012). Exploring the antecedents of logistics social responsibility: A focus on Chinese firms. *International Journal of Production Economics, 140*(1), 18–27.

Mitra, R. (2012). My country's future: A culture—Centered interrogation of corporate social responsibility in India. *Journal of Business Ethics, 106*(2), 131–147.

Morimoto, R., Ash, J., & Hope, C. (2005). Corporate social responsibility audit: From theory to practice. *Journal of Business Ethics, 62*(4), 315–325.

Morrow, D., & Rondinelli, D. (2002). Adopting corporate environmental management systems: Motivations and results of ISO 14001 and EMAS Certification. *European Management Journal, 20*(2), 159–171.

Murray, K. B., & Vogel, C. M. (1997). Using a hierarchy-of-effects approach to gauge the effectiveness of corporate social responsibility to generate goodwill toward the firm: Financial versus nonfinancial impacts. *Journal of Business Research, 38*, 141–159.

Mutch, N., & Aitken, R. (2009). Being fair and being seen to be fair: Corporate reputation and CSR partnerships. *Australasian Marketing Journal (AMJ), 17*(2), 92–98.

Nwoke, U. (2017). Corporations and development: The barriers to effective corporate social responsibility (CSR) in a neoliberal age. *International Journal of Law and Management, 59*(1), 122–146.

Orlitzky, M., Siegel, D. S., & Waldman, D. A. (2011). Strategic corporate social responsibility and environmental sustainability. *Business & Society, 50*(1), 6–27.

Parmar, B. L., Freeman, R. E., Harrison, J. S., Wicks, A. C., Purnell, L., & De Colle, S. (2010). Stakeholder theory: The state of the art. *The Academy of Management Annals, 4*(1), 403–445.

Pless, N. M. (2007). Understanding responsible leadership: Roles identity and motivational drivers. *Journal of Business Ethics, 74*(4), 437–456.

Pless, N. M., & Maak, T. (2011). Responsible leadership: Pathways to the future. In *Responsible Leadership* (pp. 3–13). Dordrecht: Springer.

Policies & Practices. (2012, September). *How to integrate adapted CSR policies in SMEs?* Retrieved from https://www.eu-ems.com/practical.asp?event_id=119&page_id=1139

Porter, M. E., & Kramer, M. R. (2006). The link between competitive advantage and corporate social responsibility. *Harvrd Business Review, 84*(12), 78–92.

———. (2007). Strategy & society: The link between competitive advantage and corporate social responsibility. *Harvard Business Review*, *84*(12), 78–81.

Prabhakar, R., & Mishra, S. (2013). A study of corporate social responsibility in Indian organization: An introspection. In *Proceedings of 21st International Business Research Conference, 10–11 June*. Toronto, Canada: Ryerson University.

Prasad, R. N. (2013). *Trusteeship, Corporate Social Responsibility and Peace*.

Kumar, P., & Ghodeswar, B. M. (2015). Factors affecting consumers' green product purchase decisions. *Marketing Intelligence & Planning, 33*(3), 330–347.

Ramus, C. A., & Steger, U. (2000). The roles of supervisory support behaviors and environmental policy in employee 'ecoinitiatives' at leading-edge European companies. *Academy of Management Journal, 43*(4), 605–626.

Rangan, K., Chase, L. A., & Karim, S. (2012). Why every company needs a CSR strategy and how to build it. Retrieved from https://hbswk.hbs.edu/item/why-every-company-needs-a-csr-strategy-and-how-to-build-it

Rassindren, M. J. G., & Hans, V. B. (2015). Corporate social responsibility—an evolutionary outline from an economic perspective. *Nitte Management Review, 9*(2), 54–69.

Reddy, N. (2018, February 12). *Want a successful business? Build an effective strategy.* Retrieved from https://www.forbes.com/sites/forbescoachescouncil/2018/02/12/want-a-successful-business-build-an-effective-strategy/#5fa8d6dd69bf

Rhodes, M. (2010, June 7). *Five essentials of an effective strategy.* Retrieved from https://managementhelp.org/blogs/strategic-planning/2010/06/07/five-essentials-of-an-effective-strategy/

Rim, H., & Song, D. (2013). The ability of corporate blog communication to enhance CSR effectiveness: The role of prior company reputation and blog responsiveness. *International Journal of Strategic Communication, 7*(3), 165–185.

Ringle, C. M., Sarstedt, M., & Straub, D. W. (2012). A critical look at the use of PLS-SEM in MIS Quarterly (EDITOR'S COMMENTS). *MIS Quarterly, 36*(1), iii–xiv.

Rolnick, P. J. (1962). Charity, trusteeship, and social change in India. *World Politics, 14*(3), 439–460.

Russo, A., & Perrini, F. (2010). Investigating stakeholder theory and social capital: CSR in large firms and SMEs. *Journal of Business Ethics, 91*(2), 207–221.

Schenider, S., & Barsoux, J. L. (2003). *Managing across cultures*. London, UK: Prentice Hall.

Seitanidi, M. M., & Crane, A. (2009). Implementing CSR through partnerships: Understanding the selection, design and institutionalisation of nonprofit-business partnerships. *Journal of Business Ethics, 85*(2), 413–429.

Sekaran, U. (2000). *Research method for business* (3rd ed.). New York, NY: John Wiley & Sons.

Sethi, S. P., Martell, T. F., & Demir, M. (2017). Enhancing the role and effectiveness of corporate social responsibility (CSR) reports: The missing element of content verification and integrity assurance. *Journal of Business Ethics, 144*(1), 59–82.

Sharma, E., & Tewari, R. (2017). Engaging employee perception for effective corporate social responsibility: Role of human resource professionals. *Global Business Review, 19*(1), 1–20.

Sharma, S. G. (2009). Corporate social responsibility in India: An overview. *International Lawyer, 43*(4), 1515.

Sivakumar, N. (2008). The business ethics of Jamsetji Nusserwanji Tata—A forerunner in promoting stakeholder welfare. *Journal of Business Ethics, 83*(2), 353–361.

Slack, K. (2012). Mission impossible? Adopting a CSR-based business model for extractive industries in developing countries. *Resources Policy, 37*, 179–184.

Smith, C. (2013). When it comes to CSR, size matters. Retrieved from https://www.forbes.com/sites/insead/2013/08/14/when-it-comes-to-csr-size-matters/#2d41fe4652a2

Stewart, H., & Gapp, R. (2014). Achieving effective sustainable management: A small-medium enterprise case study. *Corporate Social Responsibility and Environmental Management, 21*(1), 52–64.

Sundar, P. (2000). *Beyond business*. Indian Centre for Philanthropy, Tata Mcgraw Hill.

Szczanowicz, J., & Saniuk, S. (2014). Implementation of CSR concept in manufacturing SMEs. *Management, 18*(1), 71–82.

Székely, F., & Knirsch, M. (2005). Responsible leadership and corporate social responsibility: Metrics for sustainable performance. *European Management Journal, 23*(6), 628–647.

Tugrul, T. O., & Lee, E.-M. (2017). Promoting charitable donation campaigns on social media. *The Service Industries Journal*, *38*(3-4), 149–163.

United Nations Industrial Development Organization. (2002). *Corporate social responsibility: Implications for small and medium enterprises in developing countries*. Retrieved from http://www.unido.org/fileadmin/import/userfiles/puffk/corporatesocialresponsibility.pdf

Upadhyaya, R. B. (1976). *Social responsibility of business and the trusteeship theory of Mahatma Gandhi*. New Delhi, India: Sterling Publishers.

Vlachos, P. A., Panagopoulos, N. G., Bachrach, D. G., & Morgeson, F. P. (2017). The effects of managerial and employee attributions for corporate social responsibility initiatives. *Journal of Organizational Behavior, 38*, 1111–1129.

Voegtlin, C. (2011). Development of a Scale Measuring Discursive Responsible Leadership. J*ournal of Business Ethics, Special Issue on Responsible Leadership, 98*(1), 57–73, 2012; University of Zurich Chair of Foundations of Business Administration and Theories of the Firm Working Paper No. 206.

Wade, M. (2006). Developing leaders for sustainable business. In T. Maak & N.M. Pless (Eds.), *Responsible Leadership* (pp. 227–244).

Waldman, D. A. (2011). Moving forward with the concept of responsible leadership: Three caveats to guide theory and research. In *Responsible Leadership* (pp. 75–83). Dordrecht: Springer.

Waldman, D. A., & Siegel, D. (2008). Defining the socially responsible leader. *Leadership Quarterly, 19*(1), 117–131.

Waldman, D., Siegel, D. S., & Javidan, M. (2006). Components of CEO transformational leadership and corporate social responsibility. *Journal of Management Studies, 43*, 1703–1725.

Walls, J. L., Phan, P. H., & Berrone, P. (2011). Measuring environmental strategy: Construct development, reliability, and validity. *Business & Society, 50*(1), 71–115.

Walsh, J. P. (2005). Book review essay: Taking stock of stakeholder management. *Academy of Management Review, 30*(2), 426–438.

Werts, C. E., Linn, R. L., & Jöreskog, K. G. (1974). Intraclass reliability estimates: *Testing structural assumptions. Educational and Psychological measurement, 34*(1), 25–33.

Williamson, D., Lynch-Wood, G., & Ramsay, J. (2006). Drivers of environmental behaviour in manufacturing SMEs and the implications for CSR. *Journal of Business Ethics, 67*(3), 317–330.

Wood, A. J. (2017). Promoting a strategic business focus to balance competitive advantage and corporate social responsibility—Missing elements. *Social Responsibility Journal, 13*(1), 78–94.

Yoo, Y., & Alavi, M. (2001). Media and group cohesion: Relative influences on social presence, task participation, and group consensus. *MIS Quarterly, 25*(3), 371–390.

Biomass Conversion and Biorefinery
https://doi.org/10.1007/s13399-019-00596-x

ORIGINAL ARTICLE

Investigation of physicochemical properties of oil palm biomass for evaluating potential of biofuels production via pyrolysis processes

Pranshu Shrivastava[1] · Phonthip Khongphakdi[2] · Arkom Palamanit[3] · Anil Kumar[4] · Perapong Tekasakul[5]

Received: 23 September 2019 / Revised: 1 December 2019 / Accepted: 27 December 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

The objective of this work is to evaluate the potential of oil palm biomass (OPB) in terms of physicochemical properties for producing biofuels via pyrolysis processes. The OPB included oil palm trunk (OPT), oil palm fronds (OPF), oil palm shell (OPS), oil palm roots (OPR), oil palm decanter cake (OPDC), empty fruit bunches (EFB), oil palm fiber (OPFB), and oil palm sewage sludge (OPSS). Their physicochemical properties are considered on several physical, chemical, and thermal aspects. The results showed that particle size distribution and bulk density of ground OPB were different. The proximate analysis results of OPB were consistent with the lignocellulose content and extractives. The carbon and hydrogen content of the OPB were also correlated with the organic components. Some OPB contained high lignin and extractives. The lignin content of OPB strongly influenced to thermal decomposition trend. OPB contained high inorganic elements such as potassium (K), calcium (C), and iron (Fe). The higher heating value and potential use as energy equivalent with fossil fuels of the OPB were relatively low. OPB had low thermal conductivity, and the dielectric constant, loss factor, and tangent loss of the OPB were also low. Thus, these results will be beneficial for the researchers and biofuel producers for choosing the appropriate OPB, as well as the operating conditions and reactor types.

Keywords Biofuels · Biomass · Oil palm biomass · Physicochemical properties of biomass · Pyrolysis

1 Introduction

The relentless growth of the global population inevitably leads to higher energy consumption [1]. In 2018, British petroleum (BP) reported that global energy demand was 1.6 Gtoe [2]. Most of the currently available energy is obtained from fossil fuels, including crude oil, coal, and natural gas. The growth in global energy demand strongly influences to the energy systems not only in terms of energy security and sustainability but also concerning environmental and social aspects. Global warming and climate change due to greenhouse gases are already severe problems on a global scale. The grave nature of the situation has prompted both developed and developing countries around the world for searching and innovating the clean energy resources in the forms of renewable energy (RE) and alternative energy (AE) resources [1–5]. However, the potential of RE and AE resources such as solar, hydro, wind, geothermal, tidal, wave, and biomass varies significantly between countries and regions. The potential of RE and AE resources of each country and region depends on several factors such as location, weather, and season, as well as the local agricultural and industrial activities. Biomass is one of the high potential RE and AE resources in some regions or countries. Biomass is defined as organic matters which may be obtained from direct plantation (forestry), animal and human

✉ Arkom Palamanit
arkom.p@psu.ac.th

[1] Energy Technology Program, Faculty of Engineering, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand

[2] Sustainable Energy Management Program, Faculty of Environmental Management, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand

[3] Interdisciplinary Graduate School of Energy Systems, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand

[4] Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India

[5] Department of Mechanical Engineering, Faculty of Engineering, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand

Published online: 11 January 2020

waste, residues and wastes from agricultural and industrial processes, and organic municipal solid wastes [6]. In 2018, biomass used as bioenergy accounted for approximately 14% of total global energy consumption [7–10]. Bioenergy has the potential to decrease carbon footprints and to reduce environmental impacts because the replantation or new growing of plants consumes the carbon dioxide (CO_2) that is emitted from the biomass conversion processes into bioenergy [2]. For this reason, the utilization of bioenergy is considered to be carbon dioxide neutral. However, the biomass potential of each country or region depends on the location, weather, plantations, agricultural activity, and industrial processing in that specific location. Biofuels and bioenergy are most commonly produced by countries whose prime industry is agriculture. Most of the Asian countries have a high degree of dependency on the agricultural sector, with Thailand among them.

As an agro-industry country, Thailand has a high potential for biomass applications. Rice husks, cassava roots, corn stalks, rubber wood bark, rubber wood sawdust, oil palm shell, and empty fruit bunches are examples of the biomass available in Thailand [11]. These biomasses are produced during the replanting, harvesting, and processing of agricultural products. For oil palm biomass (OPB), there is particularly high potential in the south of Thailand, which is among the top-ranked producers and exporters of palm oil, after only Malaysia and Indonesia. The total plantation area of the oil palm tree in Thailand increased to 818,500 ha in 2018, which was 4.2% more than in 2017 [12]. The Thai government has also set the target of increasing the plantation area of palm oil trees to 1,632,000 ha in 2036 [12]. The OPB is obtained from harvesting and processing the fresh fruit bunches, as well as during the replantation of oil palm trees. Typically, oil palm fronds and leaves (2–3 fronds) will be produced when harvesting fresh fruit bunches from each oil palm tree [13, 14]. The replantation of oil palm trees also provides a huge source of OPB, including oil palm trunk (OPT), oil palm fronds and leaves (OPF and OPL), and oil palm roots (OPR) [14]. The milling process of fresh fruit bunches also generates various types of biomass and waste such as oil palm fiber (OPFB), empty fruit bunches (EFB), oil palm kernel shell (OPS), oil palm decanter cake (OPDC), and sewage sludge (OPSS) from wastewater treatment [15–18]. Previous reports have revealed that the yield of crude palm oil obtained from the milling process of fresh fruits was only 20–25% (wt.) [15, 17, 19]. The other parts were turned into biomass in the form of residue or waste [15–18, 20]. These reports also indicated that milling 100 tons of fresh fruit bunches generated 4–6 tons of OPS, 20–32 tons of EFB, and 12–19 tons of OPFB [14, 19–21]. The Ministry of Energy of Thailand reported that some of the OPB, such as OPS, OPFB, and EFB, has been used as solid biofuel in palm oil factories and other industries for the generation of heat and power [12]. However, OPB, including EFB, OPT, OPF, OPL, OPR, and OPDC, remains high potential for biofuel or bioenergy applications.

The conversion of biomass into biofuels or bioenergy can be performed by several processes such as mechanical conversion, thermochemical conversion, biochemical conversion, and combined processes. The criteria for choosing the biomass conversion process among the options depend on many factors such as the type and properties of the biomass being used, the desired fuel forms or energy products, the cost, storage, transportation considerations, and the utilization target. This is because the conversion of biomass by the mentioned processes provides different final forms of biofuels (liquid, gas, or solid) or bioenergy (heat or power). Pyrolysis is one of the thermochemical conversion processes. It involves the thermal decomposition of raw materials under the absence of oxygen or air at elevated temperatures ranging from 300 to 600 °C [22, 23]. This process can be applied for several materials such as plastic wastes, municipal solid wastes, waste cooking oil, used tires, and biomass. The pyrolysis of biomass provides products in the form of bio-oil, biochar, and pyrolysis gas. There are two types of pyrolysis processes – slow pyrolysis and fast pyrolysis – depending on many factors such as temperature, heating rate, and vapor residence time [24, 25]. Fast pyrolysis is more popular than slow pyrolysis for the production of bio-oil or pyrolysis oil. This is because the bio-oil yield obtained from fast pyrolysis is higher than from slow pyrolysis [26–29]. However, the yield and quality of the bio-oil or liquid product not only depend on pyrolysis types but also are influenced by many other factors such as the reactor type, operating conditions (heating rate, pyrolysis temperature, vapor residence time, etc.), biomass type, biomass property, and biomass composition [30–40]. Biomass properties such as particle size, bulk density, elemental composition, lignocellulosic composition, and chemical composition have a substantial impact on the bio-oil yield and quality [35, 40–50]. The main advantage of bio-oil is that it has a high energy density compared to raw biomass. Bio-oil is also easy to store, handle, and transport. The bio-oil also has many applications. It can be used directly as fuel in a boiler, upgraded for use as fuel in an engine, and turned into high value-added products for food and chemical processes [29, 49, 50]. For the biochar and pyrolysis gas, they can be used as biofuels to generate heat or power for use in pyrolysis or other processes.

With OPB, the applications of this biomass for pyrolysis process are still challenging due to many types of OPB that can be applied with pyrolysis processes for producing biofuels in forms of bio-oil, biochar, and pyrolysis gas. Most of the research papers related with pyrolysis have used OPS, OPFB, and EFB [14, 51–56]. There are only a few studies that have investigated the pyrolysis of other oil palm biomasses [57–59]. Thus, the literature reviews indicated that there are very less studies that revealed the properties of OPB or pyrolysis of OPB in Thailand. Most of the research papers related with OPB and pyrolysis of OPB are published by Malaysian researchers. The aim of this study was, therefore, to

investigate the physicochemical properties of OPB for evaluating the potential of biofuels production via pyrolysis processes. The OPB used in this study included OPT, OPF, OPS, OPR, OPDC, EFB, OPFB, and OPSS. The investigated physicochemical properties of these biomasses were considered via proximate analysis, ultimate analysis, higher heating value, and potential use as energy equivalent with fossil fuels, lignocellulose content, thermal decomposition via thermogravimetric analysis (TGA), elemental composition, thermal conductivity, and dielectric property.

2 Materials and methods

2.1 Preparation of biomass samples

The OPT, OPF, and OPR were obtained from oil palm trees which were harvested for 25 years old at Klong Thom District, Krabi Province, Thailand. The OPS, OPDC, EFB, OPFB, and OPSS were residue and waste from crude palm oil production at the Thiando palm oil factory located at Lam Thap District, Krabi Province, Thailand. The fresh OPT (with bark), OPF (without leaves), and OPR were chopped by a chopping machine (MCH-420, Machinery 789, Thailand). All samples were dried inside a solar greenhouse dryer to reduce the moisture content to below 10% (%wt., wet basis). The dried samples were then ground by a grinding machine (Bonny, 2HP, Thailand) equipped with a 2 mm size sieve. The ground samples were kept in a sealed plastic zip bag until ready for further use.

2.2 Determination of physicochemical properties

2.2.1 Particle size distribution

The particle size distribution of the ground OPB was determined by using a vibratory sieve shaker (Analysett 3 pro, Fritsch, Germany) comprising seven sieves with opening sizes of 0.08, 0.11, 0.15, 0.21, 0.30, 0.60, and 1.18 mm and a bottom pan (< 0.08). The determination was performed following the sieve standard ASTM E11 [47]. The result was presented as a weight percentage.

2.2.2 Bulk density

The bulk or apparent density of the ground OPB was determined based on the mass and volume ratio. The cylindrical container with a specific inside diameter and volume (500 mL) was used to determine the bulk density of the biomass samples. The ground biomass samples were poured into a container from a certain height to facilitate the free flow of the samples until the container was full with a specific volume. The net weight of each sample was then recorded, and the bulk density was calculated. To ensure the accuracy of each measurement, it was repeated three times, and the mean value with standard deviation value was presented in the unit of kg/m^3 [47].

2.2.3 Proximate and ultimate analysis

The proximate analysis is the determination of the gross components, including moisture content, volatile matter, and fixed carbon content and ash content [48]. These components were determined by using a macro thermogravimetric analyzer (TGA 701, LECO, USA), according to the ASTM D7582 procedure. The basic elemental composition of each sample, including carbon (C), hydrogen (H), oxygen (O), nitrogen (N), and sulfur (S), was determined by a CHNS/O Analyzer (FLASH 2000, Thermo Scientific, Italy). The oxygen content was determined based on the analyzer procedure. The C, H, and N were performed according to related EN15104, and the S content was performed according to the ASTM D4239. The mean and standard deviation of values were reported as percentages by weight (wt.%, dry basis).

2.2.4 Higher heating value and potential use as energy equivalent with fossil fuels

The higher heating value (HHV) of the OPB samples was determined by a bomb calorimeter (C5000, IKA® Werke, Germany), according to the EN14918 procedure. The HHV obtained from the bomb calorimeter was also compared to the HHV obtained from the calculation based on the following equations:

$$\mathrm{HHV}_1 = 0.3516\mathrm{C} + 1.16225\mathrm{H} - 0.11090 + 0.0628\mathrm{N} + 0.10465\mathrm{S}(\mathrm{MJ/kg}) \quad (1)$$

where C, H, O, and S are the elemental composition obtained from ultimate analysis ($R^2 = 0.720$) [49].

$$\mathrm{HHV}_2 = 0.3536\mathrm{FC} + 0.1559\mathrm{VM} - 0.0078\mathrm{A}(\mathrm{MJ/kg}) \quad (2)$$

where FC, VM, and A are the content of fixed carbon, volatile matter, and ash obtained from proximate analysis ($R^2 = 0.695$) [50].

$$\mathrm{HHV}_3 = 0.341\mathrm{C} + 1.322\mathrm{H} - 0.120 - 0.12\mathrm{N} + 0.068\mathrm{S} - 0.0153\mathrm{Ash}(\mathrm{MJ/kg}) \quad (3)$$

where C, H, O, S, and ash are the compositions of the biomass obtained from ultimate analysis and proximate analysis ($R^2 = 0.695$) [60].

$$\mathrm{HHV}_4 = 0.0889\mathrm{L} + 16.8218(\mathrm{MJ/kg}) \quad (4)$$

where L is the lignin content of the biomass obtained from lignocellulose content determination ($R^2 = -0.875$) [61].

The higher heating value (HHV) of the OPB obtained from the bomb calorimeter was then converted to the potential use as energy equivalent with fossil fuels, including coal (bituminous), crude oil, natural gas (NG), and liquefied petroleum gas (LPG). The reference HHV of the coal, crude oil, NG, and LPG were 28.87 MJ/kg, 38.78 MJ/L, 38.95 MJ/m^3, and 50.08 MJ/kg, respectively [62, 63].

2.2.5 Lignocellulosic content

The lignocellulosic content of the OPB was determined via the method developed by Georing and Van Soest (1970) [64], Van Soest (1991) [65], Reza et al. (2014) [66], and Kambo and Dutta (2015) [67]. According to this method, the percentage of cellulose, hemicellulose, and lignin was determined in terms of acid detergent fiber (ADF), neutral detergent fiber (NDF), and acid detergent lignin (ADL). The percentage of cellulose, hemicellulose, and lignin was then calculated from the following equations. Extractives are calculated by difference method.

$$\text{Cellulose content} = \text{ADF}-\text{ADL}(\text{wt\%, dry basis}) \quad (5)$$

$$\text{Hemicellulose content} = \text{NDF}-\text{ADF}(\text{wt.\%, dry basis}) \quad (6)$$

$$\text{Lignin content} = \text{ADL}(\text{wt.\%, dry basis}) \quad (7)$$

$$\text{Extractives} = 100-(\text{Cellulose} + \text{Hemicellulose} + \text{Lignin})(\text{wt.\%, dry basis}) \quad (8)$$

2.2.6 Thermogravimetric analysis (TGA)

Thermal decomposition of the OPB was observed by thermogravimetric analysis (TGA) and differential thermal analysis (DTA) techniques using a thermogravimetric analyzer (Perkin Elmer, USA), according to the ASTM E1131 procedure. The observation was performed at temperatures ranging from 50 to 1000 °C and at a heating rate of 10 °C/min under a nitrogen gas atmosphere.

2.2.7 Major and minor elemental compositions

The major and minor elemental contents, including silicon (Si), iron (Fe), calcium (Ca), magnesium (Mg), sodium (Na), potassium (K), cadmium (Cd), and lead (Pb), of the biomass samples were determined by the Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES, AVIO500, Perkin Elmer, USA), according to the EN15290 procedure. The results were reported as mg/kg.

2.2.8 Thermal conductivity

Thermal conductivity of biomass is one of the important factors for designing the conventional heating unit, which the heat transfers from the reaction wall to the biomass. In this study, the thermal conductivity measurement of the ground OPB was performed by the Hot Disk Thermal Constants Analyzer (Hot Disk, TPS 2500S, Sweden). This system determines the thermal conductivity based on a patented transient plane source (TPS) technique, which can be used to study materials with thermal conductivities from 0.005 to 500 W/m•K and cover a temperature range from 30 to 1000 K.

2.2.9 Dielectric properties

Dielectric properties are important factors for designing microwave heating which can be applied for pyrolysis process. Thus, this study determined the dielectric properties, including dielectric constant (ε'), dielectric loss factor (ε"), and tangent loss (tan δ) of the ground OPB by using materials test equipment (E5071C ENA Series Network Analyzer, Keysight Technologies, USA) equipped with a coaxial probe. The ground samples were poured into a cylindrical container, and then the measurement was performed at a temperature of 25 °C and a frequency of 2.45 GHz [68]. The measurement of the dielectric properties for each sample was repeated five times to gain a confidence value, and the mean value of the dielectric properties was presented.

3 Results and discussion

3.1 Particle size distribution of ground samples

Table 1 shows the particle size distribution of the ground OPB. The dried OPB samples were ground with a grinding machine equipped with a 2 mm size sieve. The determination of the particle size distribution showed that most of the ground OPB had particle sizes in the range of 0.30–1.18 mm, except the OPSS, which had a particle size of lower than 0.08 mm. For the pyrolysis process, the particle size of the biomass is one of the main factors that strongly influence product yield and quality. This is because the biomass particles with a small particle size have a high surface area, leading to a high heat transfer rate and also a high heating rate of the biomass inside the pyrolysis reactor. In the case of bio-oil production, the heating rate strongly affects the bio-oil yield and quality. With a fast pyrolysis process, the biomass is usually pyrolyzed by moving the bed pyrolysis reactors with small biomass particles. Pattiya and Suttibak [69] and Sirijanusorn et al. [36] studied bio-oil production from cassava biomass with biomass particle sizes ranging from 0 to 250 and 250–425 to 425–600 μm by using a fluidized bed and counter-rotating twin-screw reactor. They found that the maximum bio-oil yield was obtained with a particle size of 250–425 μm. Pyrolysis of biomass with particles that were too small or large tended to decrease the bio-oil yield. There are two reasons to explain

Table 1 Particle size distribution of ground biomass samples

Size		Particle size distribution of ground biomass (wt.%)							
Mesh	mm	OPT	OPF	OPS	OPR	OPDC	EFB	OPFB	OPSS
16	1.18	7.94	10.22	38.21	6.86	41.28	57.53	42.15	0.00
30	0.60	36.39	52.92	16.76	46.32	23.24	8.18	12.04	16.61
50	0.30	25.80	20.26	11.49	27.25	13.59	16.41	21.95	25.21
70	0.21	8.64	4.99	4.68	6.09	6.18	6.55	7.45	9.36
100	0.15	7.19	4.32	4.34	3.88	3.53	4.35	6.09	5.95
140	0.11	4.28	2.41	3.64	3.28	3.26	2.72	4.01	4.51
200	0.08	3.09	2.18	5.43	1.34	2.76	3.30	2.83	8.13
Pan	> 0.08	6.67	2.69	15.45	4.99	6.17	0.94	3.48	30.23
Total		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

why the biomass particle size of 0–250 μm gave a low bio-oil yield. The first reason is related to the high ash content of the biomass due to the segregation of the feed into a high dirt feed. The second reason is related to the hydrodynamic nature of the fine biomass particles. The tiny biomass particles tend to be blown or entrained out of the reaction zone before completion of the pyrolysis process. For the large biomass particles, the bio-oil yield was found to be low because the larger biomass particles have lower heat transfer. The large particles are usually not complete pyrolyzed, leading to the formation of biochar [36, 69].

The biomass particle sizes not only influence the bio-oil yield but also affect the quality of the bio-oil or liquid product in terms of water content and solid content. The water content of the bio-oil was increased in line with an increase in biomass particle size. This is because the large particles provided a low heat transfer rate. The high water content in the bio-oil also decreased its heating value [21]. For the effect of biomass particle size on solid content in the bio-oil, pyrolysis of small biomass particles using a moving bed reactor or under a high carrier gases flow rate creates the risk of producing bio-oil with high solid content. This is because the small particles have a low density, which is easily carried out from the reaction zone before the pyrolysis is complete. However, in practice, the desirable biomass particle size for pyrolysis processes may range from millimeters to centimeters, depending on the capacity and type of the pyrolysis reactors [36]. Moreover, the biomass particle size directly affects the preparation cost, transportation, storage, and feeding system [70].

3.2 Bulk density of ground samples

Figure 1 shows the bulk density of the ground OPB, which was in the range of 151.83–895.96 kg/m^3, depending on the type of OPB. The OPSS had the highest bulk density (895.96 kg/m3), followed by the OPS (702.61 kg/m^3) and the OPDC (592.88 kg/m^3), while the OPT had the lowest bulk density of 151.83 kg/m^3. When comparing the bulk density of the ground OPB with ground cassava rhizomes (230 kg/m^3) and cassava stalk (200 kg/^{3}), the results indicate that the bulk density of the OPT, OPF, OPR, and OPFB is relatively lower than that of the cassava biomass [71]. For the biofuel and bioenergy application of biomass, the bulk density of the biomass particles is one of the most important physical properties. Bulk density of biomass or biomass particles is defined as the mass of the biomass particles divided by the total volume of those particles in the standard container. The total volume includes the volume of the biomass particles, the void volume between the biomass particles, and the volume of the pores inside the biomass particles.

Usually, each biomass has a different bulk density because it has different physical features and structure, as well as different compositions. In the case of OPB particles, such as OPT, OPF, EFB, OPR, and OPPF, they have low bulk density because these biomasses are fibrous and bulky. The chopping or grinding of dried fibrous biomass into powder or small particles helps to improve its bulk density. However, the bulk density of the ground sample is relatively low compared to solid or dense particles such as in coal and sand. For the ground OPS, OPSS, and OPDC samples, they have high bulk density since these biomasses have a solid or dense structure. Grinding dense biomass provides small and dense solid particles, leading to high bulk density. In the pyrolysis process, the bulk density of the biomass particles is one of the key parameters for designing the pyrolysis reactor and feeding system of continuous pyrolysis systems. Beside the reactor size, the bulk density of biomass also directly influences energy density (MJ/m^3), storage area requirement and cost, and transportation cost [72–75].

3.3 Proximate analysis results

Table 2 shows the results of the proximate analysis, including the moisture content (MC), volatile matter (VM), fixed carbon content (FC), and ash content (AC) of the OPB samples. The results show that the MC, VM, FC, and AC of the OPB

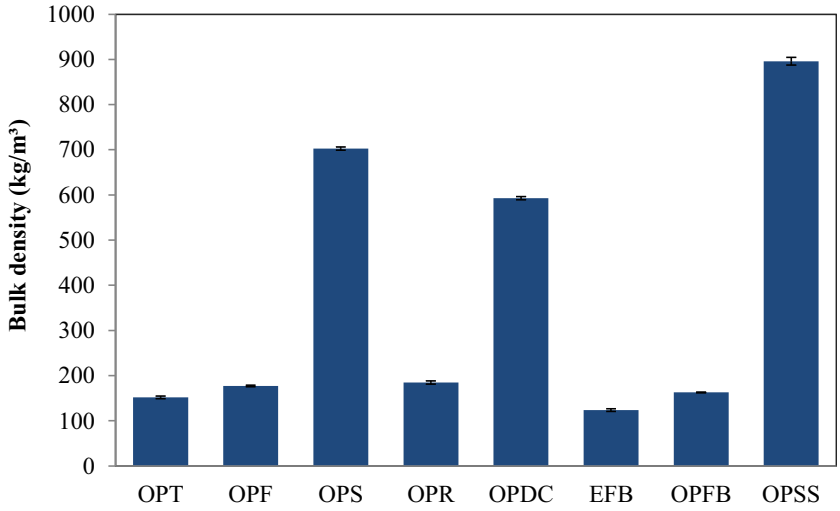


Fig. 1 Bulk density of ground oil palm biomass samples

samples were in the ranges of 8.11–14.42, 51.70–79.98, 11.42–26.14, and 1.35–36.43 (wt.%), respectively. The moisture content of the biomass samples after drying with a solar greenhouse dryer was relatively low, implying that these biomass samples are appropriate for being used as raw biomass to produce biofuels via pyrolysis processes. In pyrolysis processes, particularly in the production of bio-oil through the fast pyrolysis process, the initial moisture content of the biomass strongly influences the water content in the obtained bio-oil or liquid product [21, 76]. The water content of the bio-oil usually varies in the range of 15–40% (wt., wet basis), depending on the moisture content of the biomass and pyrolysis conditions [69, 77, 78]. The pyrolysis of biomass with high moisture content results in bio-oil with high water content. Thus, the moisture content in the biomass should be as low as possible in practice. However, the water content in the bio-oil or liquid product not only depends on the moisture content in the biomass but also is influenced by the pyrolysis conditions, such as the heating rate, reactor type, vapor residence time, and pyrolysis temperature [77, 78]. Aside from these aspects, the pyrolysis of moistened biomass also leads to a loss of energy during dehydration. In practice, the biomass can be dried to the desired moisture content lower than 10% (wt. wet basis). However, drying biomass to the low moisture content affects energy consumption and cost. Thus, the biomass can be dried from the waste heat or the heat obtained from the combustion of the pyrolysis gas.

Table 2 Proximate analysis results

Biomass	Proximate analysis results (wt.%)			
	MC	VM	FC	AC
OPT	8.21 ± 0.04	79.98 ± 0.60	16.17 ± 0.51	3.69 ± 0.05
OPF	8.11 ± 0.04	78.90 ± 0.22	17.87 ± 0.29	3.12 ± 0.05
OPS	8.92 ± 0.17	72.58 ± 0.21	26.14 ± 0.15	1.35 ± 0.53
OPR	11.92 ± 0.10	68.32 ± 0.42	22.28 ± 0.55	8.49 ± 1.05
OPDC	10.38 ± 0.05	72.13 ± 0.06	12.75 ± 0.05	14.79 ± 0.10
EFB	8.93 ± 0.04	79.16 ± 0.29	15.43 ± 0.52	5.48 ± 0.27
OPFB	9.53 ± 0.01	75.43 ± 0.56	18.24 ± 0.33	5.82 ± 0.20
OPSS	14.42 ± 0.04	51.70 ± 0.12	11.42 ± 0.10	36.43 ± 0.09

MC is in the unit of wt.% (as received basis). VM, FC, and AC are presented as wt.% (dry basis)

For volatile matter, the results reveal that the OPT, OPF, EFB, and OPFB had volatile mater in the range of 75–80% (wt.), which is close to the volatile matter results of pinewood (81% wt.) [79], cassava rhizomes (81.5% wt.,), and cassava stalk (81.2% wt.) [69]. The volatile matter of the OPR, OPS, and OPDC was in the range of 68–72% (wt.). The OPSS had the lowest amount of volatile matter (51.70% wt.), which is not much different from the works of [80–82]. The OPT, OPF, and EFB had high volatile matter content because they contain high amounts of cellulose and hemicellulose, while the amount of lignin is low (this will be discussed in more detail in the next part.). Volatile matter is an essential component of biomass for producing bio-oil via pyrolysis processes because the yield of bio-oil or liquid product depends on the volatile matter, which is the part of the solid biomass that can be converted into vapors or gases in the form of both condensable and non-condensable vapors. The previous works related to pyrolysis processes have reported that the pyrolysis of biomass with the high volatile matter provided high yields of bio-oil or liquid products [23]. In the part of fixed carbon and ash content, it can be seen in Table 2 that the OPS and OPR had fixed carbon content higher than 20% (wt.), and the OPDC and OPSS had relatively high ash content (14.79, 36.43% wt.). For the pyrolysis process, the fixed carbon content is the part of the biomass that cannot be converted into vapors or gases. This product is called biochar. Thus, for the pyrolysis of biomass to produce a high yield of biochar, biomass with the high fixed carbon content is appropriate. The last part of the proximate analysis is the ash content, which is

an inorganic composition in biomass. Biomass with the high ash content indicates that it is not suitable for bio-oil production via pyrolysis processes because it leads to a low yield of bio-oil (liquid product) and high amounts of biochar. However, producing bio-oil by the co-pyrolysis of sewage sludge with woody and non-woody biomass can help to improve the bio-oil yield [47, 80, 81].

3.4 Ultimate analysis results

Table 3 shows the ultimate analysis results, including carbon (C), nitrogen (N), hydrogen (H), sulfur (S), and oxygen (O) content of the OPB samples. The atomic ratio (H/C and O/C) of the biomass is also presented in Fig. 2. The results show that the C, N, H, O, and S content of the OPB samples were in the ranges of 32.39–53.29, 0.19–5.07, 4.83–6.92, 32.08–44.60, and < 0.01–0.77% (wt.). These results indicate that the type of OPB used affected the elemental composition. The carbon and hydrogen of the OPB samples (except OPSS) were relatively high compared to the elemental composition of the lignocellulosic biomass such as cassava rhizomes, cassava stalk, wheat straw, corn stalk, wood sawdust, corncob, and rice husk [69, 82, 83]. The results of the ultimate analysis are consistent with the proximate analysis results. High fixed carbon content in the biomass leads to a high carbon element as clearly observed in the cases of OPS, OPR, and OPFB. Sukiran (2016) [82] reported that there is a good relationship between atomic H/C and O/C ratios in OPB, indicating fairly high energy contained in the C–H and C–O bonds. Chaiwong et al. (2013) [84] explained that a high ratio of H/C indicates a high content of aliphatic hydrocarbon compounds in the biomass. A relatively low O/C ratio in the biomass indicates that it has a small number of polar compounds [84]. For the oxygen content, it was found that the OPT, OPF, EFB, and OPSS had an oxygen content of about 32–33% (wt.), which was lower than the oxygen content in the OPS, OPR, and OPFB (41.54–44.60% wt.).

For bio-oil production via pyrolysis processes, the biomass with high carbon, high hydrogen, and low oxygen content is favorable [85]. This is because carbon and hydrogen can be converted into useful aromatics, bio-oil, or liquid products. On the other hand, the oxygen will bond with the hydrocarbon molecules as phenol, acetic acid, and other oxygenated compounds, which is a disadvantage of biomass for producing bio-oil. Besides, the pyrolysis of biomass with high oxygen content presents the risk of obtaining bio-oil or liquid products with high water content. This is because reaction water can be formed from the reaction between hydrogen and oxygen [85]. The higher water content in bio-oil leads to lower heating values, as well as the instability of liquid products [76, 86]. For nitrogen and sulfur, they have a small bio-oil or liquid composition. These elements are transformed by pyrolysis into gas products. In terms of the nitrogen and sulfur content in the OPB samples, they were found to be low, except in the OPSS which had high contents of both nitrogen and sulfur. The nitrogen and sulfur content levels of the biomass samples obtained from this study are consistent with the previous studies [47, 80, 81]. The nitrogen and sulfur content in the biomass have less influence on the nitrogen and sulfur content in the bio-oil. However, most of the sulfur and nitrogen remained in the biochar [84, 87].

3.5 Minor and major elemental compositions

The minor and major elements in the OPB samples are shown in Table 4. It can be clearly seen that the minor and major elements of the OPB samples are silicon (Si), iron (Fe), calcium (Ca), magnesium (Mg), sodium (Na), and potassium (K), with the content of these elements ranging from 142.07 to 10,907.68 mg/kg for Si, 581.11 to 7959.66 mg/kg for Fe, 841.39 to 21,407.33 mg/kg for Ca, 10.98 to 18,892.55 mg/kg for Mg, 47.87 to 526.18 mg/kg for Na, and 175.63 to 70,791.31 mg/kg for K. The investigated OPB samples had very low levels of cadmium (Cd) and lead (Pb), ranging from 0.10 to 1.55 mg/kg and 0.96 to 10.47 mg/kg, respectively. The OPSS had the highest content of the minor and major elements, which is consistent with the results of the proximate analysis, indicating the highest ash content is in the OPSS. For

Table 3 Ultimate analysis results

Biomass	Ultimate analysis (wt.%, dry basis)				
	Carbon	Nitrogen	Hydrogen	Oxygen	Sulfur
OPT	45.91 ± 0.19	0.42 ± 0.00	6.38 ± 0.04	33.19 ± 0.43	< 0.01
OPF	45.94 ± 0.02	0.19 ± 0.00	6.39 ± 0.02	33.24 ± 0.09	< 0.01
OPS	53.29 ± 0.09	0.30 ± 0.01	6.15 ± 0.11	41.54 ± 0.18	0.01
OPR	50.58 ± 0.13	0.28 ± 0.00	6.28 ± 0.02	44.60 ± 0.19	<0.01
OPDC	45.60 ± 0.06	2.38 ± 0.01	6.41 ± 0.06	36.40 ± 0.37	0.24
EFB	49.73 ± 0.16	1.01 ± 0.04	6.92 ± 0.13	32.61 ± 0.35	0.07
OPFB	50.85 ± 0.05	1.05 ± 0.01	6.71 ± 0.04	42.14 ± 0.75	0.09
OPSS	32.39 ± 0.39	5.07 ± 0.09	4.83 ± 0.09	32.08 ± 0.29	0.77

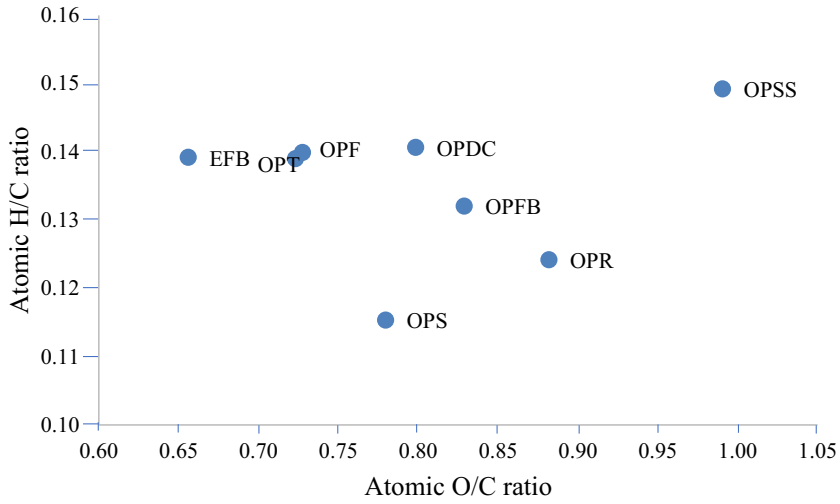


Fig. 2 Atomic ratio of hydrogen and oxygen with respect to carbon

biomass, some of the minor and major elements function as macro and micronutrients. These results indicate that OPB is a biomaterial which is safe to be used as biofuel and bioenergy because it contains very low levels of heavy metals (Cd and Pb). The obtained results were consistent with the results of Loh, 2016 [82]. However, there are many works which have revealed the alkali content of OPB, such as the amount of Na, Mg, K, and Ca, to be especially important in thermochemical conversion processes. This is because these elements may react with the Si, which then leads to operational problems in the reactors, furnaces, and boilers [82, 88, 89].

For the pyrolysis process, Alvarez et al. (2015) reported that the metals in the biomass resulted in catalytic activity on the pyrolysis products by promoting secondary reactions such as cracking and dehydration [73]. These effects led to obtaining low bio-oil yields and high gas yields. Moreover, Eom et al. (2012) concluded that some inorganic elements such as Ca, Mg, and K would promote the degradation of cellulose and hemicellulose from the biomass, favoring the formation of gas [90]. Additionally, many papers have investigated the effects of catalytic activity of inorganic elements in co-pyrolysis processes [47, 90, 91]. Zhang et al. (2015) found the synergetic effect of the co-pyrolysis of sewage sludge and rice husk [92]. The high inorganic elements in sewage sludge led to a high gas yield of co-pyrolysis. Zuo et al. (2014) also found that the co-pyrolysis of poplar sawdust and sewage sludge in a fixed bed reactor provided the maximum yield of bio-oil when they mixed 80% sludge and 20% sawdust [93]. Thus, the co-pyrolysis of OPB such as OPT, OPR, OPS, EFB, and OPFB with OPSS is still interesting.

3.6 Higher heating value and potential use as energy equivalent with fossil fuels

The higher heating value (HHV) or gross calorific value (GCV) of the OPB samples obtained from both determinations by bomb calorimeter and estimation by correlations is shown in Table 5. It can be seen that the HHV of the OPB obtained from the bomb calorimeter was in the range of 11.19–18.72 MJ/kg. The HHV (HHV_1-HHV_4) can be estimated by correlations, but the accuracy of the results depends on the correlation used. Based on the values of the estimated HHVs, they indicated that the HHV estimated by using the results from the proximate analysis provided the best

Table 4 Major and minor elements of oil palm biomass

Biomass	Major and minor elements (mg/kg, dry basis)							
	Si	Fe	Ca	Mg	Na	K	Cd	Pb
OPT	142.07	1086.96	1668.57	10.98	431.83	175.63	0.10	1.72
OPF	353.18	581.11	3817.10	1630.22	209.15	20,994.04	0.40	ND
OPS	143.71	2438.06	21,407.33	744.80	47.87	2708.62	ND	ND
OPR	426.18	7959.66	841.39	1311.93	99.01	7909.73	0.10	6.95
OPDC	364.44	1525.26	14,366.06	4737.85	96.85	27,435.65	0.19	<0.96
EFB	330.54	533.80	5103.02	2705.57	285.67	46,331.60	<0.1	ND
OPFB	375.41	889.73	4042.47	2026.11	63.02	7362.17	0.10	ND
OPSS	10,907.68	5915.44	38,886.73	18,892.55	526.18	70,791.31	<1.55	<10.47

ND means not defined

Table 5 Higher heating values of oil palm biomass

Biomass	Higher heating value (MJ/kg)				
	HHV	HHV_1	HHV_2	HHV_3	HHV_4
OPT	16.11	19.90	18.13	20.00	17.41
OPF	15.81	19.81	18.59	19.97	17.62
OPS	18.72	20.40	20.54	21.28	20.21
OPR	16.30	20.15	18.48	20.04	19.42
OPDC	17.00	19.62	15.36	19.16	18.68
EFB	18.18	21.98	17.76	21.99	17.90
OPFB	18.20	21.08	18.16	20.95	18.72
OPSS	11.19	12.77	11.81	12.48	19.53

estimation values for all biomass samples. The obtained HHV of the OPB from this study was consistent with the results obtained from Sukiran, 2016 [82]. The heating value (both LHV and HHV) of the biomass depends on many factors such the type, composition, age, plantation region, and weather conditions. The composition of the biomass is one of the most important factors that influence its heating value. Normally, biomass consists of moisture, organic components, and inorganic components.

For energy applications, the biomass should contain the lowest moisture content possible. This is because biomass with the high moisture content will release low useful thermal energy (net heating value or lower heating value). The heating value of the biomass depends on the gross components, including the volatile matter (VM), fixed carbon content (FC), and ash content (AC). Biomass with a high VM and FC, but with low AC, will provide high heating values such as in the cases of the OPS, EFB, and OPFB for OPB samples. The heating value of the biomass can also be described by ultimate analysis results based on the relative content of the main elements, including C, H, and O. This is due to the fact that the overall values of these elements are directly related to the biochemical components of the cell wall, such as the lignocellulose content [3]. In the case of OPB, it can be seen that the biomass samples with high cellulose and lignin tended to have high HHV. The previous works reported that lignin has higher levels of HHV than cellulose and starch [63]. The OPB samples have high calorific values conforming the high carbon and hydrogen content of the biomass as well as its low moisture content, as shown in Table 3.

For the pyrolysis process, the heating value of the biomass did not affect the heating value of the bio-oil. The biomass with high HHV provided bio-oil with a low HHV [69]. This is because different types of biomass with the same heating value may have different compositions (volatile matter, fixed carbon content, lignocellulose content). These components can be converted into condensable vapors under different conditions during the pyrolysis process [92]. Table 6 shows the potential use of the OPB samples as energy equivalent with fossil fuels. These HHVs were calculated to the equivalent quantity of fossil fuels, including crude oil, coal, LPG, and NG. The potential uses of the OPB samples as energy equivalent with crude oil, coal, LPG, and NG were 0.29–0.52 L, 0.39–0.70 kg, 0.22–0.40 kg, and 0.29–0.52 m^3, respectively. These results indicated that the utilization of OPB as bioenergy could help to reduce the reliance on fossil fuels and subsequently lessen environmental impacts, particularly the net carbon dioxide emission.

Table 6 Potential use as energy equivalent with fossil fuels of oil palm biomass

Biomass	HHV (MJ/kg)	Potential use as energy equivalent with fossil fuels*			
		Crude oil (L)	Coal (kg)	LPG (kg)	NG (m^3)
OPT	16.11	0.42	0.56	0.32	0.41
OPF	15.81	0.41	0.55	0.32	0.41
OPS	18.72	0.48	0.65	0.37	0.48
OPR	16.30	0.42	0.56	0.33	0.42
OPDC	17.00	0.44	0.59	0.34	0.44
EFB	18.18	0.47	0.63	0.36	0.47
OPFB	18.20	0.47	0.63	0.36	0.47
OPSS	11.19	0.29	0.39	0.22	0.29

*with respect to 1 kg of OPB

3.7 Lignocellulose content

Table 7 shows the chemical composition of the OPB samples in terms of cellulose, hemicellulose, lignin, and extractives. The cellulose, hemicellulose, lignin, and extractives of the OPB were in the range of 2.59–54.35, 8.25–25.97, 6.64–38.15, and 2.55–58.69% (wt.), respectively. The type of OPB had a strong effect on lignocellulosic compositions. The OPT, OPF, EFB, and OPFB had high cellulose and hemicellulose content. The OPS, OPR, and OPFB had high cellulose content, but their lignin content was also high. The OPDC and OPSS had high extractives, which are part of the inorganic elements and ash. The ratios of hemicellulose/lignin and cellulose/lignin are depicted in Fig. 3. It can be seen that the OPT, OPF, EFB, and OPFB had relatively high ratios, indicating high cellulose and hemicellulose content in the biomass samples. For the pyrolysis process, biomass with the high cellulose and hemicellulose contents usually produces higher yields of bio-oil or liquid than biomass with high lignin content. This is because the thermal decomposition of lignin is more difficult than that of cellulose and hemicellulose. The pyrolysis of biomass with a high lignin content usually leads to a high biochar yield [94]. The effects of the cellulose, hemicellulose, and lignin content of the biomass on the yield of pyrolysis products have been studied by many authors. Quan

Table 7 Lignocellulose content of oil palm biomass samples

Oil palm biomass	Lignocellulose content (wt.%, dry basis)			
	Cellulose	Hemicellulose	Lignin	Extractives*
OPT	39.40 ± 0.77	25.97 ± 1.17	6.64 ± 0.31	27.99 ± 0.65
OPF	54.35 ± 0.29	20.72 ± 0.12	8.96 ± 0.26	15.96 ± 0.44
OPS	37.95 ± 0.28	11.52 ± 0.01	38.15 ± 0.45	12.38 ± 0.35
OPR	49.00 ± 2.55	19.19 ± 0.06	29.26 ± 0.07	2.55 ± 0.08
OPDC	21.85 ± 0.41	13.51 ± 0.93	20.91 ± 0.32	43.73 ± 0.58
EFB	37.82 ± 0.57	21.85 ± 0.54	12.16 ± 0.17	28.17 ± 0.47
OPFB	33.58 ± 1.80	25.41 ± 1.99	21.38 ± 0.92	19.63 ± 0.86
OPSS	2.59 ± 0.81	8.25 ± 0.52	30.47 ± 0.57	58.69 ± 0.49

et al. (2016) studied the pyrolysis behavior of cellulose, hemicellulose, and lignin at 500 °C [95]. They found that the pyrolysis of these components provided liquid yields of 18.67, 30.83, and 0.5% (wt.), respectively. The mentioned effects were also investigated by Kim et al. (2013), who found a higher bio-oil yield when using biomass with higher cellulose and hemicellulose content [43]. Qu et al. (2011) explained that the pyrolysis of cellulose provided a high bio-oil yield because cellulose is more volatile than hemicellulose [96].

In addition to the liquid yield, variations in the composition of the cellulose, hemicellulose, and lignin in the biomass also influence the chemical compounds of the bio-oil or liquid products. For biomass with high levels of cellulose, the liquid fraction usually contains acids, alcohols, aldehydes, ketones, esters, heterocyclic derivatives, and phenolic compounds [97]. Stefanidis et al. (2014) studied the pyrolysis of cellulose, hemicellulose, and lignin separately [98]. They found that the liquid products obtained from cellulose pyrolysis were composed of sugars, simple phenols, ketones, aldehydes, and alcohols. The main compounds in bio-oil produced from hemicellulose pyrolysis were ketones, phenols, acids, and aldehydes. The bio-oil produced from lignin is composed of complex phenols with high molecular weights [98, 99].

3.8 Thermal decomposition behavior

The thermal decomposition behavior of the OPB via TGA and DTA is shown in Fig. 4a, b, respectively. The observation results indicate that there are four main stages (as shown in the DTA curves) in the thermal decomposition of the OPB samples. The thermal decomposition of the OPB samples also occurred at different trends due to the different components of the lignocellulosic biomass, namely, cellulose, hemicelluloses, and lignin. The first stage occurred at temperatures of 50–100 °C. At this stage, the weight of the biomass decreased slightly due to the evaporation of the moisture content. In the second stage at temperatures of 100–250 °C, the weight of the OPB was relatively constant due to the lower amount of evaporation of the light volatile compounds. Most of the thermal energy supplied to the biomass was used to increase its temperature during this stage. The next stage is the main thermal decomposition of the OPB, which was occurred at

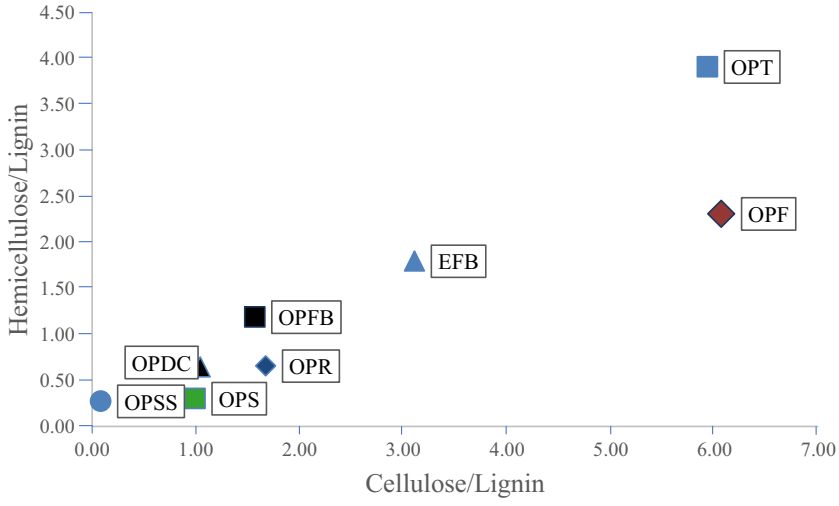


Fig. 3 The ratio of cellulose and hemicellulose, and cellulose/lignin ratio

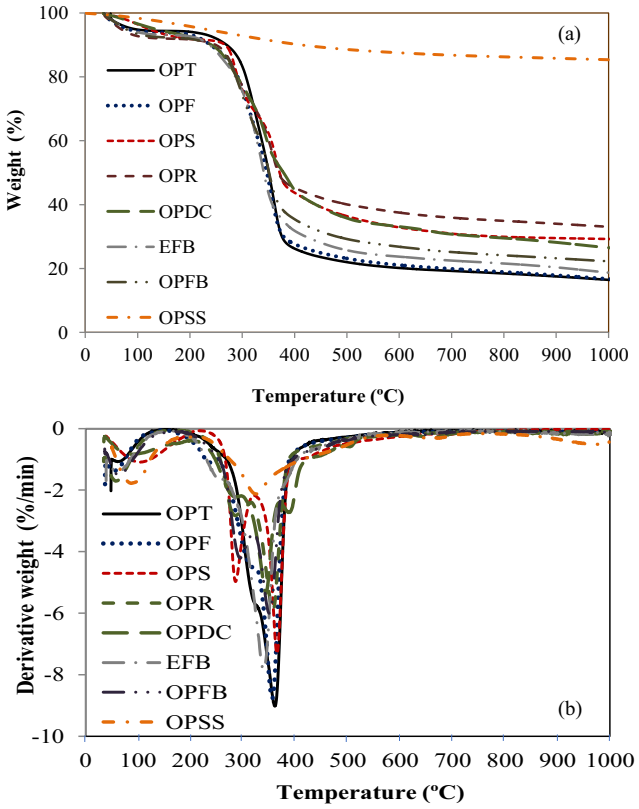


Fig. 4 TGA (a) and DTA (b) profiles of oil palm biomass samples

temperatures of 250–400 °C. For the last stage, the thermal degradation of the biomass samples was prolonged due to the decomposition of the lignin. The weight of the biomass samples decreased mainly in the third stage due to the decomposition of the cellulose and hemicellulose into both condensable and non-condensable vapors.

A previous study found that the decomposition of hemicellulose, generally represented by xylan, mainly takes place within the temperature range of 250–350 °C, followed by cellulose decomposition, which primarily occurs at the temperature range of 325–400 °C with levoglucosan as the main pyrolysis product. Recent investigations have reported that lignin was the most stable component, decomposing in the higher temperature range of 300–900 °C [100]. Figure 4a also showed the less thermal decomposition of the OPSS. This is because the main composition of the OPSS was extractives, which are mainly composed of inorganic elements and ash. The thermal decomposition trends of the OPB shown in Fig. 4a were consistent with the trends for the lignocellulosic content of the OPB samples shown in Table 7. The lignin is covalently linked to hemicellulose and cross-linked to the polysaccharide, and thus, it is highly rigid and hard. Besides, lignin consists mostly of aromatic rings; thus, the bio-oil produced from lignin biomass is rich in aromatic compounds [98].

Table 8 Thermal conductivity and dielectric properties of oil palm biomass

Oil palm biomass	Dielectric properties			Thermal conductivity (W/m•K)
	ε_r'	ε_r''	tanδ	
OPT	1.01	0.01	0.01	0.08
OPF	0.99	0.01	0.01	0.09
OPS	1.56	0.04	0.02	0.14
OPR	1.14	0.01	0.01	N/A
OPDC	0.95	0.00	0.00	0.12
EFB	1.28	0.02	0.02	N/A
OPFB	1.20	0.02	0.01	0.12
OPSS	2.03	0.10	0.05	0.17

N/A means not available

3.9 Thermal conductivity and dielectric properties

Table 8 shows the thermal conductivity and dielectric property of the OPB samples. The results reveal that the thermal conductivity of the OPB samples was in the range of 0.08–0.17 W/m•K. The OPS, OPDC, OPFB, and OPSS had thermal conductivity higher than 0.10 W/m•K, while the OPSS had the highest thermal conductivity (0.17 W/m•K). For the dielectric property of the OPB samples, it was reported in terms of dielectric constant (ε'), dielectric loss factor (ε"), and tangent loss (tan δ). The dielectric constant (ε'), dielectric loss factor (ε"), and tangent loss (tan δ) of the OPB samples were in the ranges of 0.95–2.03, 0.00–0.10, and 0.00–0.05, respectively. The dielectric property of the OPS, OPDC, OPFB, and OPSS was relatively high, which is consistent with the values of thermal conductivity. Comparing the dielectric property and thermal conductivity of the biomass samples with their inorganic elements indicated that the OPB with high Fe, K, and Ca content also provided high values of dielectric property and thermal conductivity. These results indicate that the type of OPB affected the thermal conductivity and dielectric property. The solid biomass has low thermal conductivity because it is an anisotropic material [101].

Many previous studies have determined the thermal conductivity of biomass. For example, Mason et al. (2016) reported that the thermal conductivities of dried wood pellets, miscanthus, and willow were between 0.10 and 0.12 W/m•K, while agricultural residues such as wheat and rape straws had a relatively low thermal conductivity value of 0.05 W/m•K [102]. Gupta et al. (2003) found that the thermal conductivities of softwood, softwood bark, and softwood char were in the range of 0.095–0.205 W/m•K [103]. For the pyrolysis process, the thermal conductivity of biomass strongly affects the heating systems and reactor design, heating rate, and kinetics of pyrolysis [51]. The significance of the heat transfer properties of small biomass particles in evaluating the chemical kinetics of pyrolysis and char combustion has

been described by Hayhurst [104]. Differences in thermal conductivity affect the internal temperatures and heating rates in the particle, which, in turn, affect the reaction kinetics [105].

For the dielectric property, it is the key property of the materials or biomass for microwave heating. This is because the microwave absorbability and conversion to the heat of any materials depends on the dielectric property. The ε' represents the ability of molecules to become polarized under the electric field. The ε" indicates the ability of materials to convert electromagnetic energy into heat. The ability or potential of materials that can be heated by microwave can be evaluated by the loss tangent, which is the ratio of the dielectric loss factor to the dielectric constant of the materials [52, 106]. In general, materials can be classified into three types according to their loss tangent, namely, high (> 0.5), medium (0.1–0.5), and low (< 0.1) microwave absorbing materials [107]. In the case of OPB, the loss tangent of oil palm fiber, oil palm shell, and empty fruit bunches was 0.08, 0.12, and 0.3, respectively [53, 106]. The loss tangent of biomass is low when compared to the loss tangent of carbon materials such as biochar and activated carbon (0.1–0.8). Consequently, carbon materials are good microwave absorbers with a high capacity to absorb and convert microwave energy into heat [108, 109]. Thus, the biomass pyrolysis heated by microwave irradiation could be a more promising heating technique because of the many advantages it offers over conventional heating methods [52, 106, 110].

4 Conclusions

This study determined the physicochemical properties of oil palm biomass (OPB), including oil palm trunk (OPT), oil palm fronds (OPF), oil palm shell (OPS), oil palm roots (OPR), oil palm decanter cake (OPDC), empty fruit bunches (EFB), oil palm fiber (OPFB), and oil palm sewage sludge (OPSS) to evaluate the potential use for producing biofuels via pyrolysis processes. The physicochemical properties of the OPB were considered from particle size distribution, bulk density, proximate and ultimate analysis, higher heating value and potential use as energy equivalent with fossil fuels, elemental composition, lignocellulose content, thermogravimetric analysis, thermal conductivity, and dielectric properties. The results indicate that:

- Most of the ground OPB had low bulk density, except for the OPS, OPDC, and OPSS which had relatively high bulk density.
- The OPT, OPF, OPFB, and EFB contained high amounts of volatile matter (75.43–79.98%). The OPS and OPR had high amounts of fixed carbon, while the OPDC and OPSS contained high ash content. The main elemental compositions of OPB were consistent with the proximate analysis results.
- The higher heating value (HHV) of OPB was in the range of 11.19–18.72 MJ/kg, which was relatively low as compared to fossil fuels.
- The OPF and OPR had the high cellulose content, while the OPT and OPFB had the high hemicellulose content. The OPS, OPR, and OPSS had high lignin. The extractives were high in OPDC and OPSS. The thermal decomposition trend of the OPB depended on the lignocellulosic components, particularly the lignin.
- The content of inorganic elements such as potassium (K), calcium (Ca), magnesium (Mg), and iron (Fe) in the OPB depended on the OPB types. The inorganic elements in the OPB may act as a catalyst for the pyrolysis processes, which influence the product yield and quality.
- The thermal conductivity and dielectric properties of OPB were low. The need for an appropriate design of heating unit both conventional and microwave heating. The microwave heating of the OPB may be applied with carbon materials for improving the heat generation and heating rate inside the reactor.

Therefore, based on these results, it can be concluded that some OPB has the potential to produce biofuels via pyrolysis processes, while the co-pyrolysis of the OPB may help to improve the utilization of OPB, waste, and residues. The results from this study will help the researchers and biomass industries to identify the most appropriate way to convert OPB into biofuels via pyrolysis processes.

Acknowledgments The authors express their sincere appreciation to Energy Conservation Promotion Fund, Energy Policy and Planning Office (EPPO), Ministry of Energy of Thailand.

Funding information This study was also supported by Thailand Research Fund (Contract number: MRG6280084), Prince of Songkla University (Contract number: RDO610037S), Graduate School, Prince of Songkla University, and Interdisciplinary Graduate School of Energy systems, Prince of Songkla University.

References

1. Abuelnuor AAA, Wahid MA, Hosseini SE et al (2014) Characteristics of biomass in flameless combustion: a review. Renew Sust Energ Rev 33:363–370. https://doi.org/10.1016/j.rser.2014.01.079
2. Demirbaş A (2001) Relationships between lignin contents and heating values of biomass. Energy Convers Manag 42:183–188
3. Tanger P, Field JL, Jahn CE et al (2013) Biomass for thermochemical conversion: targets and challenges. Front Plant Sci 4:218. https://doi.org/10.3389/fpls.2013.00218
4. Yelmen B, Tarık Çakir M (2016) Biomass potential of Turkey and energy production applications. Energy Sources, Part B Econ Planning, Policy 11:428–435. https://doi.org/10.1080/15567249.2011.613443

5. British petroleum (2017) BP statistical review of world energy. BP p.l.c. 2017. https://euagenda.eu/upload/publications/untitled-96772-ea/. Accessed 26 June 2019
6. Mishra RK, Mohanty K (2019) Pyrolysis characteristics, fuel properties, and compositional study of Madhuca longifolia seeds over metal oxide catalysts. Biomass Convers Biorefinery:1–17. https://doi.org/10.1007/s13399-019-00469-3
7. Abbasi T, Abbasi SA (2010) Biomass energy and the environmental impacts associated with its production and utilization. Renew Sust Energ Rev 14:919–937. https://doi.org/10.1016/j.rser.2009.11.006
8. Laser M, Lynd LR (2014) Comparative efficiency and driving range of light- and heavy-duty vehicles powered with biomass energy stored in liquid fuels or batteries. Proc Natl Acad Sci 111:3360–3364. https://doi.org/10.1073/pnas.1314039111
9. Khosravanipour Mostafazadeh A, Solomatnikova O, Drogui P, Tyagi RD (2018) A review of recent research and developments in fast pyrolysis and bio-oil upgrading. Biomass Convers Biorefinery 8:739–773. https://doi.org/10.1007/s13399-018-0320-z
10. Fernando AL, Duarte MP, Almeida J, Boléo S, Mendes B (2010) Environmental impact assessment of energy crops cultivation in Europe. Biofuels Bioprod Biorefin 4:594–604. https://doi.org/10.1002/bbb.249
11. Jagustyn B, Patyna I, Skawińska A (2013) Evaluation of physico-chemical properties of palm kernel shell as agro biomass used in the energy industry. Chemik 67:552–559
12. Preechajarn S, Prasertsri P (2018) Thailand biofuels annual. USDA foreign agricultural service. https://www.biofuelsdigest.com/bdigest/tag/thailand/thaibiofuelsannual2018. Accessed 2 July 2019
13. Hosseini SE, Wahid MA, Ganjehkaviri A (2015) An overview of renewable hydrogen production from thermochemical process of oil palm solid waste in Malaysia. Energy Convers Manag 94:415–429. https://doi.org/10.1016/j.enconman.2015.02.012
14. Asadullah M, Ab Rasid NS, Kadir SASA, Azdarpour A (2013) Production and detailed characterization of bio-oil from fast pyrolysis of palm kernel shell. Biomass Bioenergy 59:316–324. https://doi.org/10.1016/j.biombioe.2013.08.037
15. Prasertsan S, Presertsan P (1996) Biomass residues from palm oil mills in Thailand: an overview on quantity and potential usage. Biomass Bioenergy 11:387–395. https://doi.org/10.1016/S0961-9534(96)00034-7
16. Kurnia JC, Jangam SV, Akhtar S, Sasmito AP, Mujumdar AS (2016) Prolysis advances in biofuel from oil palm and palm oil processing waste: a review. Biofuel res. J. 9:332–346. https://doi.org/10.18331/BRJ2016.3.1.3
17. Loh SK (2017) The potential of the Malaysian OPB as a renewable energy source. Energy Convers Manag 141:285–298
18. Iskandar MJ, Baharum A, Anuar FH, Othaman R (2018) Palm oil industry in South East Asia and the effluent treatment technology—a review. Environ Technol Innov 9:169–185. https://doi.org/10.1016/j.eti.2017.11.003
19. Areerat K (2006) Appropriate technology evaluation for oil palm by-products utilization in Krabi province. Dissertation, Mahidol University
20. Papong S, Yuvaniyama C, Lohsomboon P, Malakul P (2004) Overview of biomass utilization in Thailand. In: ASEAN Biomass Meeting, Bangkok, pp 1–10
21. IRENA (2015) A background paper to renewable energy in manufacturing. https://www.irena.org/publications/2014/Jun/Renewable-Energy-in-Manufacturing. Accessed 2 July 2019
22. Shuit SH, Tan KT, Lee KT, Kamaruddin AH (2009) OPB as a sustainable energy source: a Malaysian case study. Energy 34:1225–1235. https://doi.org/10.1016/j.energy.2009.05.008
23. Mckendry P (2002) Energy production from biomass (part 2): conversion technologies. Bioresour Technol 83:47–54. https://doi.org/10.1016/S0960-8524(01)00119-5
24. Velghe I, Carleer R, Yperman J, Schreurs S (2011) Study of the pyrolysis of municipal solid waste for the production of valuable products. J Anal Appl Pyrolysis 92:366–375. https://doi.org/10.1016/j.jaap.2011.07.011
25. Bridgwater AV, Peacocke GVC (2000) Fast pyrolysis processes for biomass. Renew Sust Energ Rev 4:1–73
26. Yaman S (2004) Pyrolysis of biomass to produce fuels and chemical feedstocks. Energy Convers Manag 45:651–671. https://doi.org/10.1016/S0196-8904(03)00177-8
27. Şensöz S, Can M (2002) Pyrolysis of pine (Pinus Brutia ten.) chips: 1. effect of pyrolysis temperature and heating rate on the product yields. Energy Sources 24:347–355. https://doi.org/10.1080/00908310252888727
28. Shastri Y, Hansen A, Rodríguez L, Ting KC (2014) Engineering and science of biomass feedstock production and provision. Springer Science + Business media, New York
29. Lozano-García B, Parras-Alcántara L (2013) Short-term effects of olive mill by-products on soil organic carbon, total N, C: N ratio and stratification ratios in a Mediterranean olive grove. Agric Ecosyst Environ 165:68–73
30. Bridgwater AV (1999) Principles and practice of biomass fast pyrolysis processes for liquids. J Anal Appl Pyrolysis 51:3–22
31. Meier D, Faix, O (1999) State of the art of applied fast pyrolysis of lignocellulosic materials-a review. Bioresour Technol 68, 71–77. https://doi.org/10.1016/S0960-8524(98)00086-8
32. Bridgwater AV (2012) Review of fast pyrolysis of biomass and product upgrading. Biomass Bioenergy 38:68–94. https://doi.org/10.1016/j.biombioe.2011.01.048
33. Isahak WNRW, Hisham MWM, Yarmo MA, Yun Hin T (2012) A review on bio-oil production from biomass by using pyrolysis method. Renew Sust Energ Rev 16:5910–5923. https://doi.org/10.1016/j.rser.2012.05.039
34. Fonts I, Gea G, Azuara M, Ábrego J, Arauzo J (2012) Sewage sludge pyrolysis for liquid production: a review. Renew Sust Energ Rev 16:2781–2805. https://doi.org/10.1016/j.rser.2012.02.070
35. Sirijanusorn S, Sriprateep K, Pattiya A (2013) Pyrolysis of cassava rhizome in a counter-rotating twin screw reactor unit. Bioresour Technol 139:343–348
36. Alvarez J, Amutio M, Lopez G, Bilbao J, Olazar M (2015) Fast co-pyrolysis of sewage sludge and lignocellulosic biomass in a conical spouted bed reactor. Fuel 159:810–818
37. Huang X, Cao J-P, Zhao X-Y et al (2016) Pyrolysis kinetics of soybean straw using thermogravimetric analysis. Fuel 169:93–98. https://doi.org/10.1016/j.fuel.2015.12.011
38. Wang Y, Zeng Z, Tian X, Dai L, Jiang L, Zhang S, Wu Q, Wen P, Fu G, Liu Y, Ruan R (2018) Production of bio-oil from agricultural waste by using a continuous fast microwave pyrolysis system. Bioresour Technol 269:162–168
39. Leng L, Huang H (2018) An overview of the effect of pyrolysis process parameters on biochar stability. Bioresour Technol 270:627–642
40. Chen X, Li S, Liu Z, Chen Y, Yang H, Wang X, Che Q, Chen W, Chen H (2019) Pyrolysis characteristics of lignocellulosic biomass components in the presence of CaO. Bioresour Technol 287:21493
41. Pattiya A (2011) Bio-oil production via fast pyrolysis of biomass residues from cassava plants in a fluidised-bed reactor. Bioresour Technol 102:1959–1967
42. Pattiya A, Suttibak S (2012a) Influence of a glass wool hot vapour filter on yields and properties of bio-oil derived from rapid pyrolysis of paddy residues. Bioresour Technol 116:107–113

43. Kim SS, Ly HV, Kim J, Choi JH, Woo HC (2013) Thermogravimetric characteristics and pyrolysis kinetics of alga Sagarssum sp. biomass. Bioresour Technol 139:242–248
44. Kelkar S, Saffron CM, Chai L et al (2015) Pyrolysis of spent coffee grounds using a screw-conveyor reactor. Fuel process Technol 137:170–178. https://doi.org/10.1016/j.fuproc.2015.04.006
45. Biswas B, Pandey N, Bisht Y, Singh R, Kumar J, Bhaskar T (2017) Pyrolysis of agricultural biomass residues: comparative study of corn cob, wheat straw, rice straw and rice husk. Bioresour Technol 237:57–63
46. Fiore S, Berruti F, Briens C (2018) Investigation of innovative and conventional pyrolysis of ligneous and herbaceous biomasses for biochar production. Biomass Bioenergy 119:381–391. https://doi.org/10.1016/j.biombioe.2018.10.010
47. Ani FN, Zailani R (1997) Characteristics of pyrolysis oil and char from oil Palm shells. Dev. Thermochem. Biomass convers. Springer Netherlands, Dordrecht, pp 425–432. https://doi.org/10.1007/978-94-009-1559-6_33
48. Okoroigwe E, Li Z, Stuecken T, Saffron C, Onyegegbu S (2012) Pyrolysis of Gmelina arborea wood for bio-oil/bio-char production: physical and chemical characterisation of products. J Appl Sci. https://doi.org/10.3923/jas.2012.369.374
49. Saidur R, Abdelaziz EA, Demirbas A, Hossain MS, Mekhilef S (2011) A review on biomass as a fuel for boilers. Renew. Sustain. Energy rev. 15:2262–2289. https://doi.org/10.1016/j.rser.2011.02.015
50. Acda MN (2016) Physical and chemical properties of fuel pellets from agricultural residues. Former Philipp Agric 99:283–287
51. Abdullah NA, Mohamed R, Mahmood WMFW, Saad MHM (2017) Black smoke elimination via PID controlled co-firing technique at palm oil mill. Int J Appl Eng Res 12:8050–8056
52. Salema AA, Ani FN (2011) Microwave induced pyrolysis of OPB. Bioresour Technol 102:3388–3395
53. Omar R, Idris A, Yunus R, Khalid K, Isma MIA (2011) Characterization of empty fruit bunch for microwave-assisted pyrolysis. Fuel 90:1536–1544
54. Ruengvilairat P, Tanatavikorn H, Vitidsant T (2012) Bio-oil production by pyrolysis of oil palm empty fruit bunch in nitrogen and steam atmospheres. J Sustain Bioenergy Syst 2:75
55. Kabir G, Mohd Din AT, Hameed BH (2017) Pyrolysis of oil palm mesocarp fiber and palm frond in a slow-heating fixed-bed reactor: a comparative study. Bioresour Technol 241:563–572
56. Sukiran MA, Loh SK, Bakar NA (2016) Production of bio-oil from fast pyrolysis of oil palm biomass using fluidised bed reactor. J Energy Technol Policy 6:52–62
57. Yakub MI, Abdalla AY, Feroz KK, Suzana Y, Ibraheem A, Chin SA (2015) Pyrolysis of oil palm residues in a fixed bed tubular reactor. Journal of Power and Energy Engineering 3:185–193
58. Sareekam N, Kamarudin SK, Kasmuri NH (2016) Optimization of bio oil from palm oil fronds via fast pyrolysis. Indian J Sci Technol 9:1–13
59. Bensidhom G, Hassen-Trabelsia AB, Alper K, Sghairoun M, Zaafouri K, Trabelsi I (2018) Pyrolysis of date palm waste in a fixed-bed reactor: characterization of pyrolytic products. Bioresour Technol 247:363–369
60. Annamalai K, Sweeten JM, Ramalingam SC (1987) Technical notes: estimation of gross heating values of biomass fuels. Trans ASAE 30:1205–1208
61. Parikh J, Channiwala S, Ghosal G (2005) A correlation for calculating HHV from proximate analysis of solid fuels. Fuel 84:487–494. https://doi.org/10.1016/j.fuel.2004.10.010
62. Sheng C, Azevedo JLT (2005) Estimating the higher heating value of biomass fuels from basic analysis data. Biomass Bioenergy 28:499–507
63. Demirbaş A (2001) Relationships between lignin contents and heating values of biomass. Energy Convers Manag 42:183–188
64. Goering HK, Van Soest PJ (1970) Forage fiber analysis, apparatus, reagents, procedures, and some applications. In: Agriculture handbook, vol. 379. ARS-USDA, Washington, DC
65. Van Soest PJ, Robertson JB, Lewis BA (1991) Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. J Dairy Sci 74:3583–3597
66. Reza MT, Uddin MH, Lynam JG, Hoekman SK, Coronella CJ (2014) Hydrothermal carbonization of loblolly pine: reaction chemistry and water balance. Biomass Convers Biorefinery 4:311–321. https://doi.org/10.1007/s13399-014-0115-9
67. Kambo HS, Dutta A (2015) Comparative evaluation of torrefaction and hydrothermal carbonization of lignocellulosic biomass for the production of solid biofuel. Energy Convers Manag 105:746–755. https://doi.org/10.1016/j.enconman.2015.08.031
68. Salema AA, Yeow YK, Ishaque K et al (2013) Dielectric properties and microwave heating of oil palm biomass and biochar. Ind Crop Prod 50:366–374. https://doi.org/10.1016/j.indcrop.2013.08.007
69. Pattiya A, Suttibak S (2012) Production of bio-oil via fast pyrolysis of agricultural residues from cassava plantations in a fluidised-bed reactor with a hot vapour filtration unit. J Anal Appl Pyrolysis 95:227–235
70. Duong TL, Nguyen DT, Nguyen HHM, Phan BMQ, Nguyen HL, Huynh TM (2019) Fast pyrolysis of Vietnamese waste biomass: relationship between biomass composition, reaction conditions, and pyrolysis products, and a strategy to use a biomass mixture as feedstock for bio-oil production. J Mater Cycles Waste Manag 21:624–632
71. Sirijanusorn S, Sriprateep K, Pattiya A (2013) Pyrolysis of cassava rhizome in a counter-rotating twin screw reactor unit. Bioresour Technol 139:343–348. https://doi.org/10.1016/j.biortech.2013.04.024
72. Scott DS, Piskorz J (1984) The continuous flash pyrolysis of biomass. Can J Chem Eng 62:404–412. https://doi.org/10.1002/cjce.5450620319
73. McKendry P (2002) Energy production from biomass (part 1): overview of biomass. Bioresour Technol 83:37–46
74. Brar JS, Singh K, Wang J, Kumar S (2012) Cogasification of coal and biomass: a review. Int J For Res 2012:1–10. https://doi.org/10.1155/2012/363058
75. Elliott DC, Hart TR, Neuenschwander GG, Rotness LJ, Olarte MV, Zacher AH (2012) Catalytic hydroprocessing of fast pyrolysis bio-oil from pine sawdust. Energy and Fuels 26:3891–3896. https://doi.org/10.1021/ef3004587
76. Okoroigwe EC, Saffron CM (2012) Determination of bio-energy potential of palm kernel shell by physicochemical characterization. Niger J Technol 31:329–335
77. Kang B-S, Lee KH, Park HJ, Park Y-K, Kim J-S (2006) Fast pyrolysis of radiata pine in a bench scale plant with a fluidized bed: influence of a char separation system and reaction conditions on the production of bio-oil. J Anal Appl Pyrolysis 76:32–37
78. Lu Q, Li W-Z, Zhu X-F (2009) Overview of fuel properties of biomass fast pyrolysis oils. Energy Convers Manag 50:1376–1383
79. Amutio M, Lopez G, Alvarez J, Olazar M, Bilbao J (2015) Fast pyrolysis of eucalyptus waste in a conical spouted bed reactor. Bioresour Technol 194:225–232
80. Inguanzo M, Domınguez A, Menéndez JA, Blanco CG, Pis JJ (2002) On the pyrolysis of sewage sludge: the influence of pyrolysis conditions on solid, liquid and gas fractions. J Anal Appl Pyrolysis 63:209–222
81. Domínguez A, Menéndez JA, Inguanzo M, Pis JJ (2006) Production of bio-fuels by high temperature pyrolysis of sewage

sludge using conventional and microwave heating. Bioresour Technol 97:1185–1193
82. Sukiran MA, Loh SK, Bakar NA (2016) Production of bio-oil from fast pyrolysis of OPB using fluidised bed reactor. J Energy Technol Policy 6:52–62
83. Maddi B, Viamajala S, Varanasi S (2011) Comparative study of pyrolysis of algal biomass from natural lake blooms with lignocellulosic biomass. Bioresour Technol 102:11018–11026
84. Chaiwong K, Kiatsiriroat T, Vorayos N, Thararax C (2013) Study of bio-oil and bio-char production from algae by slow pyrolysis. Biomass Bioenergy 56:600–606
85. Pimenidou P, Dupont V (2012) Characterisation of palm empty fruit bunch (PEFB) and pinewood bio-oils and kinetics of their thermal degradation. Bioresour Technol 109:198–205
86. De Bari I, Barisano D, Cardinale M, Matera D, Nanna F, Viggiano D (2000) Air gasification of biomass in a downdraft fixed bed: a comparative study of the inorganic and organic products distribution. Energy Fuel 14:889–898. https://doi.org/10.1021/ef990243g
87. Berndes G, Börjesson P, Ostwald M, Palm M (2008) Multifunctional biomass production systems–an overview with presentation of specific applications in India and Sweden. Biofuels, Bioprod Biorefining Innov a Sustain Econ 2:16–25
88. Wei X, Schnell U, Hein K (2005) Behaviour of gaseous chlorine and alkali metals during biomass thermal utilisation. Fuel 84:841–848. https://doi.org/10.1016/j.fuel.2004.11.022
89. Ninduangdee P, Kuprianov VI (2014) Combustion of palm kernel shell in a fluidized bed: optimization of biomass particle size and operating conditions. Energy Convers Manag 85:800–808
90. Eom I-Y, Kim J-Y, Kim T-S, Lee S-M, Choi D, Choi I-G (2012) Effect of essential inorganic metals on primary thermal degradation of lignocellulosic biomass. Bioresour Technol 104:687–694
91. Zabeti M, Nguyen TS, Lefferts L, Heeres HJ, Seshan K (2012) In situ catalytic pyrolysis of lignocellulose using alkali-modified amorphous silica alumina. Bioresour Technol 118:374–381
92. Zhang W, Yuan C, Xu J, Yang X (2015) Beneficial synergetic effect on gas production during co-pyrolysis of sewage sludge and biomass in a vacuum reactor. Bioresour Technol 183:255–258
93. Zuo W, Jin B, Huang Y, Sun Y (2014) Characterization of top phase oil obtained from co-pyrolysis of sewage sludge and poplar sawdust. Environ Sci Pollut Res 21:9717–9726
94. Akhtar J, Amin NS (2012) A review on operating parameters for optimum liquid oil yield in biomass pyrolysis. Renew Sust Energ Rev 16:5101–5109
95. Quan C, Gao N, Song Q (2016) Pyrolysis of biomass components in a TGA and a fixed-bed reactor: thermochemical behaviors, kinetics, and product characterization. J Anal Appl Pyrolysis 121:84–92
96. Qu X, Liang P, Wang Z, Zhang R, Sun D, Gong X (2011) Pilot development of polygeneration process of circulating fluidized bed combustion combined with coal pyrolysis. Chem Eng Technol 34:61–68
97. Klass DL (1988) The US biofuels industry. Institute of Gas Technology, Chicago
98. Stefanidis SD, Kalogiannis KG, Iliopoulou EF, Michailof CM, Pilavachi PA, Lappas AA (2014) A study of lignocellulosic biomass pyrolysis via the pyrolysis of cellulose, hemicellulose and lignin. J Anal Appl Pyrolysis 105:143–150
99. Chang G, Miao P, Yan X, Wang G, Guo Q (2018) Phenol preparation from catalytic pyrolysis of palm kernel shell at low temperatures. Bioresour Technol 253:214–219
100. Dhyani V, Bhaskar T (2018) Chapter 2: kinetic analysis of biomass pyrolysis. In: Waste biorefinery. Elsevier, London
101. Cai J, He Y, Yu X, Banks SW, Yang Y, Zhang X (2017) Review of physicochemical properties and analytical characterization of lignocellulosic biomass. Renew Sust Energ Rev 76:309–322
102. Mason PE, Darvell LI, Jones JM, Williams A (2016) Comparative study of the thermal conductivity of solid biomass fuels. Energy Fuel 30:2158–2163
103. Gupta M, Yang J, Roy C (2003) Specific heat and thermal conductivity of softwood bark and softwood char particles☆. Fuel 82:919–927
104. Hayhurst AN (2013) The kinetics of the pyrolysis or devolatilisation of sewage sludge and other solid fuels. Combust Flame 160:138–144
105. Jones JM, Saddawi A, Dooley B, Mitchell EJS, Werner J, Waldron DJ (2015) Low temperature ignition of biomass. Fuel Process Technol 134:372–377
106. Huang Y-F, Chiueh P-T, Lo S-L (2016) A review on microwave pyrolysis of lignocellulosic biomass. Sustain Environ Res 26:103–109
107. Yin C (2012) Microwave-assisted pyrolysis of biomass for liquid biofuels production. Bioresour Technol 120:273–284
108. Lam SS, Chase HA (2012) A review on waste to energy processes using microwave pyrolysis. Energies 5:4209–4232
109. Menéndez JA, Arenillas A, Fidalgo B, Fernández Y, Zubizarreta L, Calvo EG (2010) Microwave heating processes involving carbon materials. Fuel Process Technol 91:1–8. https://doi.org/10.1016/j.fuproc.2009.08.021
110. Robinson J, Dodds C, Stavrinides A, Kingman S, Katrib J, Wu Z (2015) Microwave pyrolysis of biomass: control of process parameters for high pyrolysis oil yields and enhanced oil quality. Energy Fuel 29:1701–1709

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Multi View Image Registration and Removal of Transformation Errors

S Kumar,
MAIT, Sector 22, Rohini,
Delhi 86, India,
+91-9711188513,
s_mathurec@yahoo.co.in

R Gupta
MAIT, Sector 22, Rohini,
Delhi 86, India,
+91- 9811379465,
rashmig71@yahoo.com

R Kapoor,
DCE, Bawana,
Delhi 45, India.
+91- 9999432549,
Raj_himani@hotmail.com

A Bhattacharyya
DCE, Bawana, Delhi 45,
India.
+919899191994
asok_bhattacharya@yahoo.com

ABSTRACT
This method is based upon the image registration process and the application is when the text which is to be identified is behind the mesh which works as a hurdle. We know that the mesh as hurdle can be made less irritating by either moving the camera or the source itself. The method uses Radon Transform for extracting the mesh lines and capturing the position of the mesh lines. The final process of filling the deformed image is through the registration. The method is adaptive to movement in any direction. Quaternion method is used to remove the scaling and rotational errors. It was tested on a number of images approximately and gave excellent results.

Categories and Subject Descriptors
I.4. Image Processing and Computer Vision: 1.4.3 Enhancement

Keywords:
Image Registration, Image Processing, Image in motion, Quaternion

1. INTRODUCTION
In computer vision, set of data acquired by sampling the same scene or object at different times, or from different perspectives, will be in different coordinate systems. Image registration is the process of transforming the different sets of data into one coordinate system. Registration is necessary in order to be able to compare or integrate the data obtained from different measurements. The Multi View Image Registration technique uses images of an object taken from different views called modalities.

Many interesting and useful approaches to the multiview and Medioni, and Masuda and Yokoya both incrementally register views against a growing global union of view points [20], [4]. Bergevin et al.place all views into a global frame of reference, and then repeatedly select a view and register it against all others [2]. Stoddart and Hilton find pairwise correspondences between points in all views, and then iteratively minimize correspondence errors over all views using a descent algorithm [12].This basic technique is extended using a multiresolution framework, surface normal processing, and boundary point processing by Neugebauer [19] and Eggert et al. [7]. Williams and Bennamoun suggested a further refinement by including individual covariance weights for each point [10]. Sawhney et al. and Shum and Szeliski perform the global alignment of two dimensional mosaics by nonlinear minimization of distance between point correspondences [3], [6]. Benjemaa and Schmitt propose a nonlinear solution based on the quaternion representation [18]. Their formulation is a multiview extension of the pairwise solution proposed by Horn [17] using distance between pairwise correspondences as the optimization criterion. Donghun Lee, Jihun Park [8] presents a quaternion-based method to estimate the camera parameters in a single video stream to interpolate 3D trajectory of the moving camera with camera parameters. Jihun Park [16] determining camera parameters for three dimensional scene reconstruction given an image with three known points on a single plane on the image .Robert Bergevin, Marc Soucy, Hewe Gagnon, and Denis Laurendeau [17] present an algorithm that reduces the level of the registration errors between all pairs in a set of range views. Chang Yuan, Ge´ rard Medioni, Jinman Kang, and Isaac Cohen [10] present a method for detecting motion regions in video sequences observed by a moving camera in the presence of a strong parallax due to static 3D structures. The proposed method classifies each image pixel into planar background, parallax, or motion regions by sequentially applying 2D planar homographies, the epipolar constraint, and a novel geometric constraint called the “structure consistency constraint.” Jung-Young Son, Vladimir V. Saveljev, Jai-Soon Kim, Kae-Dal Kwack, and Sung-Kyu Kim [9] Proposed a method based on 4*4 homogeneous matrices which can predict distortions in the perceived image in projection-type multiview 3-D imaging systems. Jiangbo Lu, Hua Cai, Jian-Guang Lou, and Jiang Li [5] propose a DE technique to accelerate the disparity search by employing epipolar geometry.

In this paper, the proposed algorithm is a practical approach for solving the global registration parameters using the relative motion between view pairs as the error criterion. This criterion does not require linearization and, therefore, can be used even when the accumulated rotational error is large. Furthermore, this criterion does not require point correspondences and can therefore be used together with robot odometery or any registration method. These different modalities are then analyzed and the information obtained

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ICWET’10, February 26–27, 2010, Mumbai, Maharashtra, India.
Copyright 2010 ACM 978-1-60558-812-4…$10.00.

RIGHTSLINK

is fused to form a coherent composition called multimodal image. This paper addresses the problem of large scale multiview registration of range images captured from unknown viewing directions. The global error distribution method described in can be used together with any pair wise registration algorithm to perform globally consistent multi view range image registration. The paper is concerned with the existence of hindrance in the view of an object, like when a mesh appears in between the camera and the object thus making it partially unrecognizable. The hindrance in the view or the mesh can be considered as a kind of periodic noise and the technique multi view image registration has been used for the solution of this problem. The application of image registration is useful if and only if either mesh, camera or the object is moving relative to other two i.e. each frame in unique in terms of information with respect to the previous and the next frame. So the three possible situations are:

(i) Mesh and the object are stationary but camera is in motion.

(ii) Camera and the object are stationary but mesh is in motion

(iii) Camera and mesh are stationary but object is in motion.

Here, the first situation has been considered in which mesh and the object are stationary but the camera is moving in different directions i.e. at different angles to collect the necessary information about the object which is hidden in the view at some angle but is accessible at a different angle. The movie is then converted into frames. One frame is taken as a reference frame and the rest are target frames. Radon transform to detect hindrance in the view i.e. mesh in the reference frame and the fusion technique, Pixel based fusion to fill new information collected from target frames in place of hindrance. The method is an extract of the natural registration process and hence the approach is proven. The paper is arranged as follows: Section-2 gives a briefing of the method followed. Section-3 explains the pre-processing, which in this case is the separation of the frames and the removal of the wired mesh. Section-4 explains feature-extraction and the Section-5 explains the registration algorithm. The next Section-6 gives the results obtained followed by the conclusions in Section-7.

2. MULTIVIEW EGISTRATION PROCESS

Multiview video or free-viewpoint video is an exciting application, because it enables users to watch a static or dynamic scene from different viewing angles. Generally, to provide a smooth multi-perspective viewing experience, content producers need to capture a distinct scene with ideal quality from multiple camera positions, such as the multiview camera setup shown in Fig. 1, where a multiview image set is produced by steadily moving a single video camera around a scene of interest along a predefined capture trajectory The simultaneous multiple video streams from multiview cameras are referred to as multiview video. A multiview video sequence can be regarded as a temporal sequence of special-visual-effect snapshots, captured from different viewpoints at multiple times. Multiview image set can be generated by multiple cameras which are positioned at different angles to capture the scene from different angles. Multiview image set can also be generated when the camera is stationary whereas the target image is moving. Here in this paper we consider that a single camera is moving while the target is stationary.

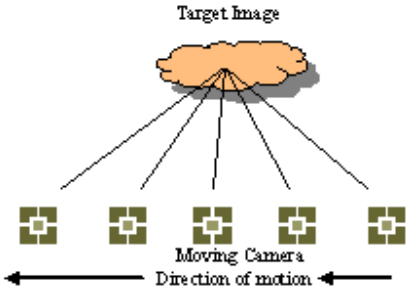


Figure1. Multiview detection scheme

2.1 Error due to Transformations

In Figs. A, the x and y axes are shown as correctly aligned and mutually perpendicular. In Figure 2 the x_1 and y_1 pair of axes are shown as rotated through an angle α relative to the x, y pair. The x_1, and y_1 pair are still perpendicular, but as a coordinate frame it is not correctly aligned. This is called drift or rotational error.

In addition to the drift or rotational error, there is a second kind of possible distortion of a coordinate frame called scale error. Pure scale error exists when the x_1 and y_1 axes, respectively, coincide with the x and y axes, but the vectors which are supposed to be unit vectors fail to have unit length.

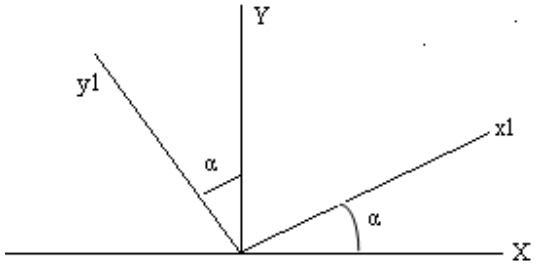


Figure 2. Rotation of X and Y axes

2.1.1 Rotational Error

D =

$$\begin{vmatrix} qw^2+iqx^2-jqy^2-kqz^2 & 2*qx*qy-2*qz*qw & 2*qx*qz+2*qy*qw \\ 2*qx*qy+2*qz*qw & qw^2-iqx^2+qy^2-kqz^2 & 2*qy*qz-2*qx*qw \\ 2*qx*qz-2*qy*qw & 2*qy*qz+2*qx*qw & qw^2-iqx^2-qy^2+kqz^2 \end{vmatrix} \quad (1)$$

Let l,m and n be the components of angular velocity as functions of time. Then q(t) is found as a solution of the system of differential equations

$$\begin{vmatrix} qw' \\ qx' \\ qy' \\ qz' \end{vmatrix} = 1/2 \begin{vmatrix} 0 & -l & -m & -n \\ l & 0 & n & -m \\ m & -n & 0 & l \\ n & m & -l & 0 \end{vmatrix} \begin{vmatrix} qw \\ qx \\ qy \\ qz \end{vmatrix} \quad (2)$$

Let's denote the square matrix above as W then we can rewrite

$$q' = \frac{1}{2} Wq \quad (3)$$

RIGHTSLINK

Let the transition matrix corresponding this be Z (t, t_o).Then we may represent the solution of (2) as

$$q(t) = Z(t, t_o)q_0 \tag{4}$$

Where, q_0 represents the initial conditions of quaternion at time t_0. Now suppose that f represents an additive noise perturbation to (4) due to error. Let the corresponding perturbed solution be p(t). We will assume that the initial value, $p(t_0)$, is likewise equal to q_0. The perturbed differential equation is

$$p' = (1/2)Wp + f \tag{5}$$

The solution to this equation is

$$p(t) = Z(t, t_0)q_0 + \int_{t_0}^{t} Z(t,\tau)f(\tau)d\tau \tag{6}$$

Denote the integral term in (6) by r (t). Comparing (4) and (6)

$$p(t) = q(t) + r(t) \tag{7}$$

By definition,

$$\| q \| = 1 \tag{8}$$

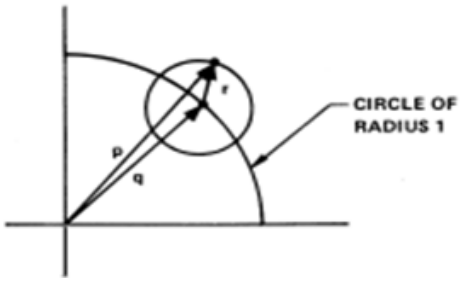


Figure 3. Quarterion Perturbation

A geometrical analog in two dimensions for the quaternion perturbation is shown in Figure 3. The quaternion q is always confined to the circumference of a circle of radius 1. The perturbation r may be in any direction with equal likelihood. Given only the vector p, without being able to make reference to q, the only way we can tell that a perturbation has occurred will be from the fact that

$$\| q \| \neq 1 \tag{9}$$

Since there is no way we can estimate the angular deviation of p relative to q, the only realistic correction we can make is to rescale p back to the circumference of the unit circle in such a way that the angular deviation is equally likely to be positive or negative. Therefore, introduce the corrected quaternion p* obtained by

$$p^*(t) = p(t)/\|p(t)\| \tag{10}$$

$$\|p(t)\| = [p_1^2(t) + p_2^2(t) + p_3^2(t) + p_4^2(t)]^{1/2} \tag{11}$$

Since the true quaternion q is not available to us, use p* in place of q in (1) in order to compute D. Now construct a 4*4 matrix S having the quaternion components as its elements

$$S = \begin{vmatrix} wq & xq & yq & zq \\ -xq & wq & zq & -yq \\ -yq & -zq & wq & xq \\ -zq & yq & xq & wq \end{vmatrix} \tag{12}$$

Also introduce a related matrix T as

$$T = \begin{vmatrix} wq & -xq & -yq & -zq \\ xq & wq & zq & -yq \\ yq & -zq & wq & xq \\ zq & yq & -xq & wq \end{vmatrix} \tag{13}$$

Considering D as an orthogonal direction cosine matrix, the transformation matrix in four dimensional spaces can be written as

$$A_4 = \begin{vmatrix} 1 & 0_{1*3} \\ 0_{3*1} & D \end{vmatrix} \tag{14}$$

$$A_4 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & d11 & d12 & d13 \\ 0 & d21 & d22 & d23 \\ 0 & d31 & d32 & d33 \end{vmatrix} \tag{15}$$

These matrices commute. Further by quaternion properties we can say

$$A = ST = TS \tag{16}$$

Now apply the above procedure to an error analysis. By (7), the effect of computational errors will be that the matrixes S and T are formed from the elements of p rather than q. Denote the results as S_p and T_p. Further, denote the matrixes formed in the same way using the components of r by S_r and T_r.

$$S_r = \begin{vmatrix} r1 & r2 & r3 & r4 \\ -r2 & r1 & r4 & -r3 \\ -r3 & -r4 & r1 & r2 \\ -r4 & r3 & -r2 & r1 \end{vmatrix} \tag{17}$$

$$T_r = \begin{vmatrix} r1 & -r2 & -r3 & -r4 \\ r2 & r1 & r4 & -r3 \\ r3 & -r4 & r1 & r2 \\ r4 & r3 & -r2 & r1 \end{vmatrix} \tag{18}$$

Then considering (7)

$$S_p = S + S_r$$

$$T_p = T + T_r \tag{19}$$

From (15)

$$S_pT_p = (S+S_r)(T+T_r) = ST + S_rT + ST_r + S_rT_r \tag{20}$$

Neglecting the term S_rT_r Direct computation will reveal that the pair of matrixes S_r and T, as well as the pair of matrixes S and T_r, actually commute. Rearranging (20)

$$S_pT_p = ST(I_4 + S'S_r + T'T_r) \tag{21}$$

Here S' and T' are the transpose of matrixes S and T. Let the perturbed version of A be called B. Then equation (21) may be written as

$$B_4 = A_4 (I_4 + H_4) \tag{22}$$

Here I_4 is a 4*4 identity matrix and

RIGHTSLINK

$$H_4 = S'S_r + T'T_r \quad (23)$$

the matrix H4 may be split into its symmetric part $\hat{S}_4$ and its anti symmetric part U_4:

$$\hat{S}_4 = \frac{1}{2}(H_4 + H'_4) \quad (24)$$

$$U_4 = \frac{1}{2}(H_4 - H'_4) \quad (25)$$

The matrix $\hat{S}_4$ represents the scale and skew errors, and the matrix U_4 represents the drift errors, which have crept into the computation of rotation via quaternion.

Reconsidering the equation (5). The sign of the perturbation vector f could, in fact, may be plus or minus. Therefore, the sign of r in (7) could equally likely be plus or minus. Similarly the sign of H_4 in (23) could equally likely be plus or minus. Hence, the algebraic signs associated with $\hat{S}_4$ and U4 may be plus or minus. Substituting (12), (13) and (17), (18) into equation (23) and the result then putting in (24) one finds that

$$\hat{S}_4 = 2(q.wr_1 + q.xr_2 + q.yr_3 + q.zr_4)\, I_4 \quad (26)$$

Therefore, if the elements of $\hat{S}_4$ are positive, it means, referring back to Fig. B that the tip of r lays outside the unit circle. Similarly, if $\hat{S}_4$ has negative elements, it means that the tip of r lies within the unit circle. At the same time, the associated rotational error could equally likely be in either direction.

We can now readily see that our proposed correctional procedure, represented by (10), will have the effect of reducing $\hat{S}_4$ to zero. Furthermore, while this procedure will not drive U_4 to zero, it will, in fact, almost always reduce the norm of U4. At worst, in a few situations (i.e., when IIPII = 1 to begin with), the correctional procedure will have no effect on U_4.

3. THE PRE-PROCESSING

The first step is to extract the frames from the video. One of these frames is chosen to be the reference frame and the rest are target frames. Since the images have been taken from a single optical device, they are classified as single-modality images. The reference image is processed further so that hindrance is removed from the view and the target images are registered into the reference image to make the object recognizable. Following is the method followed:

Algorithm 1: Extraction of frames

1. Read in the video file
2. Determine the suitable number of frames required per second
3. Extract the frames from the video clip
4. Choose an appropriate reference image from these frames

Let the captured views lie within coordinate frames V1,...,Vn, where two frames Vi and Vj are related by the rotation and translation (Ri,j, ti,j). A point pj in frame Vj can be described in frame Vi according to the change of coordinates

$$pi = Ri,j\ pj + ti,j \quad (27)$$

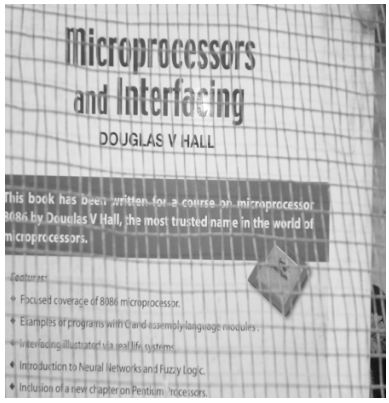


Figure 4(a). Reference frame (original)

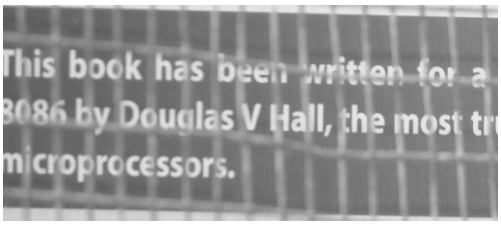


Figure 4(b). Target frame (original)

In practice, the change of coordinates between neighboring views will be found using odometery or a pair wise registration procedure [18]. We shall call such changes of coordinates, measurements.

Algorithm 2: Mesh Detection

Given: frames

1. for all available frames, do
2. Read in the grayscale image IMG
3. RADON←Radon_Transform (IMG)
4. for ANGLE← (1 to 3) & (88 to 92) & (178 to 180), do
   [to consider only the vertical and horizontal mesh lines]
5. If RADON[I, ANGLE] >= Mean(Peak_values)
6. SORTED_RADON[I, ANGLE]←RADON(I, ANGLE)
7. end if
8. end for
9. INV_RADON←inverse_Radon_Transform(SORTED_RADON)
10. [WIDTH,START_X,START_Y]←Mesh_Stats(INV_RADON)
11. MAPPED_IMG← Mapping (IMG,HOR_GAP,VERT_GAP, START_X,START_Y,WIDTH)
12. end for

The next step is to detect and remove the mesh lines. In order to determine the sections of the pictures to be fused, the exact coordinates of the pixels corresponding to the noise in the picture

RIGHTSLINK

should be known. The noise in the reference image is in the form of mesh which is nothing but an array of horizontal and vertical lines. Since, the radon transformation is a very powerful technique for the detection of straight lines; it has been used for the purpose of mesh detection. The detection of the mesh lines is achieved with the Algorithm-2

1. First the image is read as a three dimensional matrix with each element representing some intensity values Corresponding to three colors red, green and blue (RGB image).

2. Before applying radon transform, the image is converted into grayscale image which is a two dimensional image with intensity values varying from 0 to 255, 0 corresponding to black and 255 corresponding to white.

3. Radon transform is applied on the grayscale image to detect the lines representing the mesh. Figure-6(a) shows the radon transform for figure-5. Figure-6(b) shows the sorted out peak values from the Radon transform that correspond to the straight of the mesh.

The Radon transform gives projection of the image intensity along a radial line oriented at a specific angle varying form 00 to 1800. Radon transform provides with the radial coordinates of the lines corresponding to each projection and corresponding intensity values.

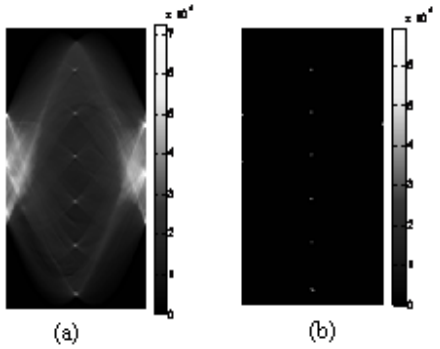


Figure 5. Grey scale image

The Radon transform for a set of parameters (rho, theta) is the line integral through the image g(x, y), where the line is positioned corresponding to the value of (rho, theta). The delta () is the Dirac delta function which is infinite for argument 0 and zero for all other arguments (it integrates to one) [15].

Equation (28) and (29) represent the above expressions.

$$\check{g}(\rho,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\delta(\rho - x\cos\theta - y\sin\theta)\,dx\,dy \qquad (28)$$

And the identical expression

$$\check{g}(\rho,\theta) = \int_{-\infty}^{\infty} g(\rho\cos\theta - s\sin\theta, \rho\sin\theta + s\cos\theta)\,ds \qquad (29)$$

The very strong property of the Radon transform is the ability to extract lines (curves in general) from very noisy images as shown below in Figure-6.

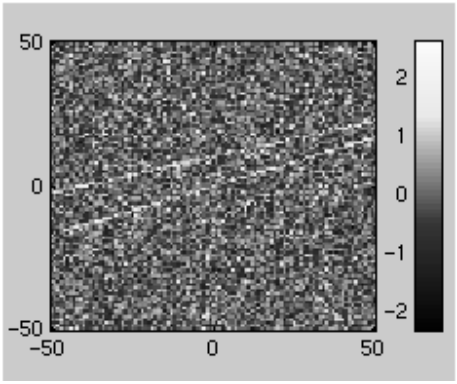


Figure 6 (a). Randon Transformation of Figure 5

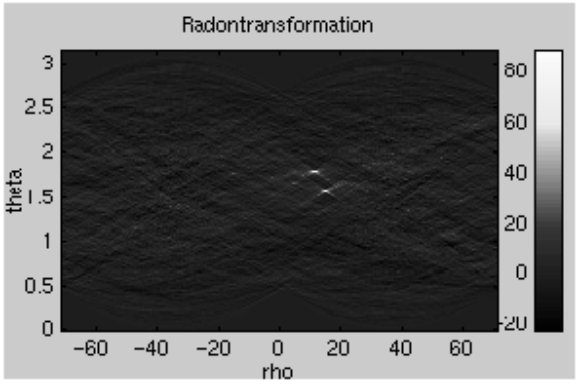


Figure 6 (b). Sorted peak values from Randon Transformation

From the Radon transform, shown in Figure-7, it can be seen that crossing lines makes no problem.

The peak points of intensity values correspond to straight lines in the reference image. A simple thresholding algorithm is applied to pick out the line parameters.

4. Inverse radon transform is applied to the line parameters obtained in the previous step which give lines representing the mesh. The inverse radon transform inverts the radon transform, and can therefore be used to reconstruct images from the projection data. It uses the radial coordinates of the line and the angles of projection to get the lines in the image. Radon uses parallel-beam projections to reconstruct the image. In parallel-beam geometry, each projection is formed by combining a set of line integrals through an image at a specific angle. As the number of projections i.e. length of theta increase, the reconstructed image is more accurate so the length of theta vector has been taken from 00 to 1800 and theta must increase monotonically with a constant increment of 10.

5. These lines are further used to map the mesh or the noise and find the total area along with the coordinates of pixels where the information is to be filled during fusion.

RIGHTSLINK

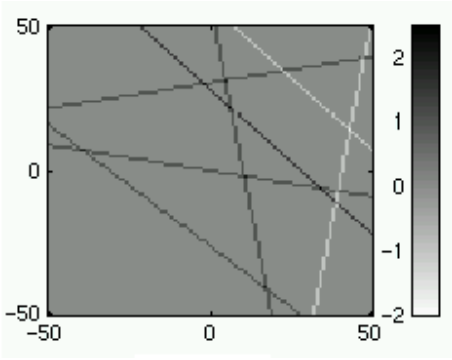


Figure 7 (a). Randon Transformation showing the lines

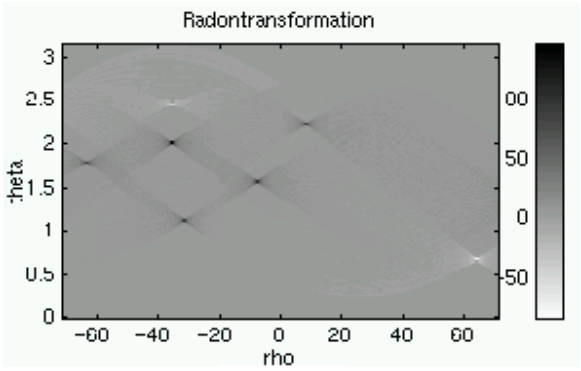


Figure 7 (b). Removal of lines by Randon Transformation

4. FEATURE EXTRACTION

In order to obtain the new information to be filled following parameters should be known.

1. Quantization of Information in Successive Frames

The amount of the information changes in successive frames should not exceed the width of mesh i.e. should not change abruptly, because if it does the embedded information will overlap the previous information and the final image will be distorted. So the determination of optimized number of frames captured per second is very important before they are processed.

2. Direction of change in the Information

Since a camera has been used for the movie of the object, the camera can be considered as a point that can move randomly in all possible directions at any possible angle in the space but only in a fix plane. The direction of movement of camera can be classified in two major categories.

i. Perfect horizontal or vertical motion.

ii. The inclined motion

The direction vector is calculated by dividing the information in between the mesh of the reference image into 4, 9or 16 equal parts depending upon the need of accuracy or computational time. More is the number of parts, higher is the accuracy but more will be the computational time. So, need is to find out the optimized value, which comes out to be 9. The information from the next frame from the same coordinates is calculated and divided into 9 equal parts. Now the correlation of each corresponding part is calculated. The part which has undergone maximum change gives the lowest value for correlation. Thus the minimum value of correlation gives an approximate value of direction vector i.e. movement of camera. The direction vector 2, 4, 6 and 8 correspond to perfect horizontal and vertical motion and 1, 3, 7 and 9 designate the inclined motion respectively. (Figure- 8 shows a single cell from the target frame Figure 4).

3. Pixel Position of New Information

The registration process being implemented is feature-based registration so the feature changed in the new frame with respective to the previous one have to be extracted before the fusion is done. The quantization determines the amount of features changed and direction vector determines the coordinates of new information i.e. the new information is specific for a particular detection vector and extraction of features is also divided into two categories:

First is when the information is changing linearly exactly horizontally or vertically. The new information is rectangular and can be easily detected by merely comparing the new frame with the previous one along each column.

Figure 8. Single cell from the target frame Figure 4

Second condition deals with the condition when the camera is moving at an inclination i.e. information enters into the frame diagonally. This time new information is not uniform and extraction is much more complicated and is drawn out by successively comparing the new frame with the older one along both rows and columns.

5. FUSION TECHNIQUE

The fusion technique being used is pixel based fusion and involves placement of information obtained from previous step in place of mesh. The coordinates of the mesh in the reference frame have already been determined like the previous step fusion also depends upon the direction vector The new information is embedded into the reference frame which gives noise free information i.e. views of object without any kind of hindrance. (Figure-9 shows the resultant image).

Figure 9. The resultant image

RIGHTSLINK

6. REGISTRATION ALGORITHM

The Following Algorithm (Algorithm-3) is the one used to implement the multi view registration of the images.
The purpose of the algorithm is to perform the pixel based fusion on a pair of frames. One is the reference frame called the base frame (BASE_IMG) and other is one of the target frames chosen for fusion from the video clip. The basic approach taken is to first divide the whole picture into elementary rectangular cells/blocks of the mesh, determine the direction of motion, and calculate the drift of the pixel positions to determine the picture section to be fused and finally fuse the calculated pixel positions.

Algorithm 3: Image Fusion

Given BASE_IMG, NEXT_FRAME

1. Read in the actual reference image ORIGINAL_BASE

2. Divide BASE_IMG into elementary rectangular mesh cells such that no part of the mesh **lines is included**

3. Divide the NEXT_FRAME into cells with same coordinates as of the BASE_IMAGE cells

4. for all cells in the BASE_IMG, do

5. BASE_CELL←Current BASE_IMG cell

6.NEXT_FRAME_CELL←Current NEXT_FRAME cell

7. Divide both cells in 9 equal partitions

8. for all corresponding partition pairs in both cells, do

9. CORR_VAL[6]←correlation(BASE_PART,NEXT_PART)

10. end for

11. DIRECTION←Max(CORR_VAL[ ])

12. H_DRIFT←Find_horizontal_drift(NEXT_FRAME_CELL)

13. V_DRIFT←Find_vertical_drift(NEXT_FRAME_CELL)

14. for Drifts in both orientations, do

15. From the edge of cell, move DRIFT pixels away along the length of edge

16. Fuse this calculated section of selected pixels from the NEXT_FRAME into BASE_IMG

17. end for

18. end for

The algorithm stated above reads in a base frame, BASE_IMG along with a target frame from the clip, NEXT_FRAME. BASE_IMG is then divided into the cells of the mesh. The same coordinates of division are then followed to divide NEXT_FRAME image. Next a loop is started that repeats the further processing on each of these cells. This processing involves, dividing corresponding cells of both the images into nine equal parts. The partitions obtained are then compared for both the images using the correlation. The partition pair with minimum value for correlation indicates the direction of the motion. After this, the Horizontal and vertical drifts of pixel positions from BASE_IMG to NEXT_FRAME are calculated in accordance with the direction of the motion. With these steps all the necessary data for fusion is available. This data gives the section of the image NEXT_FRAME to be fused with the BASE_IMG. Fusion is performed to fill up the mesh lines in BASE_IMG with the information extracted from the NEXT_FRAME that gives the resultant image free of noise i.e. the mesh.

7. RESULTS AND DISCUSSION

This paper shows the practical implementation of the registration process that the human brain does automatically when some one looks at an object through the mesh and by moving the head or the mesh, he is able to recognize the object completely. The results from the applied method have been quite encouraging. The extent of the distortion in the results is observed to be sufficiently low. The data for the application were taken under the assumptions that the camera motion involved only horizontal and vertical translations without any rotations, that the motion is confined to a single plane and that the lighting conditions and background view do not change considerably. Although the rotational variances could be nullified by applying opposite rotational transformations on the extracted frames, the process was avoided in the samples for the sake of simplicity.

8. CONCLUSION

The method was tried on 200 images having wire mesh as on the fore ground. This method removed all of them. The method works very well till the background does not change but need modification for accommodating the change in the background. It has been realized that the background change due to shadow only disturbs the registration process. This can be overcome by re-registration process but it is time consuming. The application is new and the method is insensitive to any kind of skew occurring due to the movement of camera in unknown direction.

9. THE REFERENCES

[1] A.J. Stoddart and A. Hilton, “Registration of Multiple Point Sets”, Proc. Int’l Conf. Pattern Recognition, p. B6A.5, 1996.

[2] B.K.P. Horn, “Closed Form Solutions of Absolute Orientation using Unit Quaternion”, J. Optical Soc. of Am.-A, vol. 4, no. 4, pp. 629-642, Apr.1987.

[3] Chang Yuan, Ge´ rard Medioni, Jinman Kang, and Isaac Cohen “Detecting Motion Regions in the Presence of a Strong Parallax from a Moving Camera by Multiview Geometric Constraints” IEEE Trans. Pattern Analysis and Machine Intelligence, vol... 29, NO. 9, September 2007.

[4] D.W. Eggert, A.W. Fitzgibbon, and R.B. Fisher, “Simultaneous Registration of Multiple Range Views for Use in Reverse Engineering of CAD Models”, Computer Vision and Image Understanding, vol. 69, no. 3, pp. 253-272, Mar. 1998.

[5] Donghun Lee, Jihun Park “Estimation of Camera Parameters from a Single Moving Camera using Quaternion-based Interpolation of 3D Trajectory” Computer Graphics, Imaging and Visualisation (CGIV 2007) IEEE2007.

[6] Gregory C. Sharp Sang W. Lee, and David K. Wehe, “Multiview Registration of 3D Scenes by Minimizing Error between Coordinate Frames”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 26, NO. 8, pp. 1037 to 1050 August 2004.

RIGHTSLINK

[7] H.S. Sawhney, S. Hsu, and R. Kumar, "Robust Video Mosaicing through Topology Inference and Local to Global Alignment," Proc. European Conf. Computer Vision, pp. 103-119, 1998.

[8] H.Y. Shum and R. Szeliski, "Systems and Experiment Paper: Construction of Panoramic Image Mosaics with Global and Local Alignment," Int'l J.Computer Vision, vol. 36, no. 2, pp. 101-130, Feb.2000.

[9] J.-G. Lou, H. Cai, and J. Li, "A real-time interactive multi-view video system," in Proc. ACM Int. Conf. Multimedia, Singapore, Nov. 2005, pp. 161–170.

[10] J. Williams and M. Bennamoun, "A Multiple View 3D Registration Algorithm with Statistical Error Modeling," IEICE Trans. Information and Systems, vol. 83, no. 8, pp. 1662-1670, Aug. 2000.

[11] Jiangbo Lu, Hua Cai, Jian-Guang Lou, and Jiang Li "An Epipolar Geometry-Based Fast Disparity Estimation Algorithm for Multiview Image and Video Coding" IEEE Transaction on Circuit And System For Video Technology, Vol.17, No. 6, June2007.

[12] Jihun Park "Quaternion-Based Camera Calibration and 3D Scene Reconstruction" Computer Graphics, Imaging and Visualisation (CGIV 2007), IEEE2007.

[13] Jung-Young Son, Vladimir V. Saveljev, Jai-Soon Kim, Kae-Dal Kwack, and Sung-Kyu Kim "Multiview Image Acquisition and Projection" Journal of Display Technology, Vol. 2, No. 4, December 2006.

[14] P.J. Neugebauer, "Reconstruction of Real-World Objects via Simultaneous Registration and Robust Combination of Multiple Range Images," Int'l J.Shape Modeling, vol. 3, nos. 1-2, pp. 71- 90, 1997.

[15] Peter Toft, "The Radon Transform Theory and Implementation", IMM, DTU, 1996.

[16] R. Benjemaa and F. Schmitt, "A Solution for the Registration of Multiple 3-D Point Sets Using Unit Quaternions", Proc. European Conf. Computer Vision, 1998.

[17] R. Bergevin, M. Soucy, H. Gagnon, and D. Laurendeau, "Towards a General Multi-View Registration Technique," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 5, 540- 547, May, 1996.

[18] Robert Bergevin, Marc Soucy, Hewe Gagnon, and Denis Laurendeau "Towards a General Multi-view Registration Technique" IEEE Transaction on pattern Analysis and Machine Intelligence, Vol-18 May 1996.

[19] T. Masuda and N. Yokoya, "A Robust Method for Registration and Segmentation of Multiple Range Images," Computer Vision and Image Understanding, vol. 61, no. 3, pp. 295-307, May 1995.

[20] Y. Chen and G.G. Medioni, "Object Modeling by Registration of Multiple Range Images," Image and Vision Computing, vol.0, no. 3, pp. 145-155, 1992.

RIGHTSLINK

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/337158924

N gram Based Machine Translation for English Assamese: Two Languages with High Syntactical Dissimilarity

Article · December 2019

DOI: 10.35940/ijeat.B2320.129219

CITATIONS
0

READS
16

3 authors, including:

Zakir Hussain
National Institute of Technology, Silchar
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Abdul Hannan
Gauhati University
7 PUBLICATIONS **8** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Statistical Machine Translation System for English-Assamese using ngram model View project

Identification of Assamese and Bodo Language View project

All content following this page was uploaded by Zakir Hussain on 30 December 2019.

The user has requested enhancement of the downloaded file.

N-gram based Machine Translation for English-Assamese: Two Languages with High Syntactical Dissimilarity

Zakir Hussain, Malaya Dutta Borah, Abdul Hannan

Abstract: ***To bridge the language constraint of the people residing in northeastern region of India, machine translation system is a necessity. Large number of people in this region cannot access many services due to the language incomprehensibility. Among several languages spoken, Assamese is one of the major languages used in northeast India. Machine translation for Assamese language is limited compared to other languages. As a result, large number of people using Assamese language cannot avail lots of benefits associated with it. This paper has focused on the development of the English to Assamese translation system using n-gram model. The n-gram model works very well with the language pair having high dissimilarity in syntax compared to other models. The value of n has a very big role in the quality and efficiency of the system. Bilingual Evaluation Understudy (BLEU) score differs significantly with the change of the n-gram. This model uses tuples to reduce the consumption of excess memory and to accelerate the translation process. Parallel corpus has been used for training the n-gram based decoder called MARIE. The number of translation units extracted using n-gram model is much less than the translation units extracted using phrase based model. This has a high impact on system efficiency.***

Keywords: Statistical Machine Translation, N-gram, MARIE, English-Assamese Translation, Tuple Extraction

I. INTRODUCTION

Machine translation is a subfield of computer linguistics that deals with the translation carried out by the computers. In the context of text translation, user inputs texts of one standard language and the system translates the same into the texts of another standard language. During this process, the system needs to follow some rules specifically applicable to the target language. Since the syntax of different languages is different, the system should understand the rules of different languages. The position of noun, verb and object of different languages are different; hence, it should be taken care of such that the sentences are translated with proper meaning without violation of grammatical rules. The quality of the translation is measured on the basis of some factors like handling the linguistic typology, translation of the idioms and how the anomalies are isolated.

Now-a-days, machine translation has got much importance in different fields. Earlier, language was a very big barrier in sharing ideas and knowledge between diverse language speakers. Machine translation has attracted various domain of work due to its ability of bridging the gap created by that barrier. The significant demand for translation of electronic text on the internet, such as web pages, e-mail, social media, electronic chat, official document translation etc. is being noticed from past few years. To fulfill the need of translation, the following well defined approaches are referred by the research community in the area of machine translation (Please refer to Fig.1 for pictorial view of the approaches).

- **Direct approach**: The direct approach is the simplest one. Translation is done word-by-word basis. In this approach, no linguistic analysis of the source sentence is taken into consideration for producing a target sentence. Now-a-days this approach has been abandoned even in the corpus-based framework.
- **Rule based approach**: This approach is based on the rules generated by the human experts [11], [17]. Different human experts may specify different rules for translation process. Hence, for different persons, the system will be of different configuration and of different efficiency. This approach can be subdivided into:
  - *Transfer approach*: The transfer approach has three phases. First one is the *analysis phase* where analysis of the source sentence is done to produce an abstract representation. Second one is the *transfer phase* where abstract representation of the first phase is being transferred into the equivalent representation in the target language. Third one is the *generation phase* where target sentence is generated from the intermediary representation.
  - *Interlingua approach*: This approach produces the target sentence based on the Interlingua representation created by thorough analysis of the syntax and semantics of the source sentence. This approach analyses the source sentence deeply. The advantage of Interlingua approach is that once the meaning of the source sentence is grasped, it can be articulated in any number of target languages.
- **Corpus based approach**: This type of system extracts the knowledge by analyzing the translation examples from parallel corpus. A parallel corpus is a collection of texts of more than one language, each of which is an exact translation of each other. The parallel corpus is developed by human experts. Here, the translation system can be developed as soon as the required technique is ready for a given pair of languages. A corpus-based approach generally follows direct or transfer approach. The Corpus based approach can be further subdivided into the following two categories:

Revised Manuscript Received on December 08, 2019
***** Correspondence Author
Zakir Hussain*, Research Scholar, Department of CSE, NIT Silchar, Assam, India.
Malaya Dutta Borah, Assistant Professor, Department of CSE, NIT Silchar, Assam, India.
Abdul Hannan, Faculty, Department of IT, Gauhati University, Assam, India.

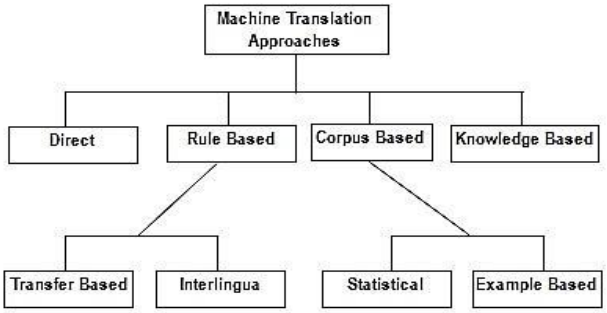


Fig. 1. Approaches to Machine Translation

- *Statistical*: It uses the examples from the parallel corpus for the training of the system. Statistical approach initially worked only on word-by-word fashion. But now-a-days, it aims to introduce linguistic analysis to some extent.
- *Example based*: It uses the examples from the parallel corpora. Translation is provided by choosing and combining the examples.

- **Knowledge based**: This approach requires *ontological knowledge* and *lexical knowledge*. Ontological knowledge means the description of the concepts and relationships that can exist for the agents and lexical knowledge means all the information that is known about the words and the relationships among them [5].

In recent research, rule based and corpus based approaches have got importance as these approaches have shown significant improvement in the quality of the translation. Due to the high syntactical dissimilarity between English and Assamese languages, the researchers have found the corpus based approach as the suitable one. To accomplish the translation, different models extract translation units in different way (please refer to section 3 for details). In the earlier works on English to Assamese translation, the researchers extracted the translation units as the phrases. That worked well but the following limitations have been found:

- The extracted translation units (phrases) are more in number and hence consuming more memory.
- Finding the desired translated words from more number of units resulted more time to translate.

To address these issues, this work proposes an *n-gram based approach* which will certainly improve the capability of the translation system in terms of the *memory consumption* as well as the *translation time*.

The focal points of this approach are:

- Reduction in the number of extracted translation units (tuples) without compromising the quality of translation
- Reduction in the translation time.

In this work we are focusing only on English to Assamese translation because we have found that for other languages, a significant level of works in the field of machine translation has already been carried out and the people across the globe is availing the benefits. But the work is lagging behind for Assamese language and hence a very big number of people cannot avail lots of benefits associated with it.

The remaining portion of this paper is arranged into 5 sections. Section 2 is about the related works, section 3 presents the proposed work, section 4 shows the experimental setup and section 5 is about the results and discussions, while section 6 concludes the paper.

II. RELATED STUDY

Machine translation system was constrained being used in early days and could be seen uniquely in some areas like military administrations, national and international legislative administrations. For translating Russian military scientific and technical documentations into English, the US Air Force introduced Systran in 1970 [21]. Later, the translation system got importance in different fields and many works have been carried out for various languages. Only a few of those have been mentioned here. Using n-gram model, Spanish-English, Arabic-English and Chinese-English translation system was developed in the TALP Research Center under Speech Processing Group in the Signal Theory and Communications department of Universitat Polit`ecnica de Catalunya. MARIE, the n-gram based machine translation decoder was also been developed there [5]. English to South Dravidian translation system was developed at Computational Engineering and Networking Department, Amrita Vishwa Vidyapeetham, Coimbatore, Tamil Nadu, India [23]. English to Sanskrit translation system has also been developed using lexical parser at Information Technology Department, Bharati Vidyapeeth University College of Engineering, Pune- 43, Maharashtra, India [1]. Translation system for English to Hindi was developed at Thapar University [20]. Likewise many more can be mentioned for different languages other than Assamese language. It has been seen that many popular translation systems like Google Translate, Bing Translator etc. have not worked for the Assamese language (Fig.2). During past a few years, the development of English-Assamese translation system has come into the interest of research community. Using phrase based decoder, translation system for English-Assamese and vice versa has been developed at Gauhati University [3] as well as at Dibrugarh University [19]. Both the systems have been implemented using very small corpus and MOSES as the decoder. MOSES is a phrase based machine translation decoder. In Gauhati University, a work using rule based machine translation approach for English-Assamese has also been done [9]. Till date limited works have been carried out

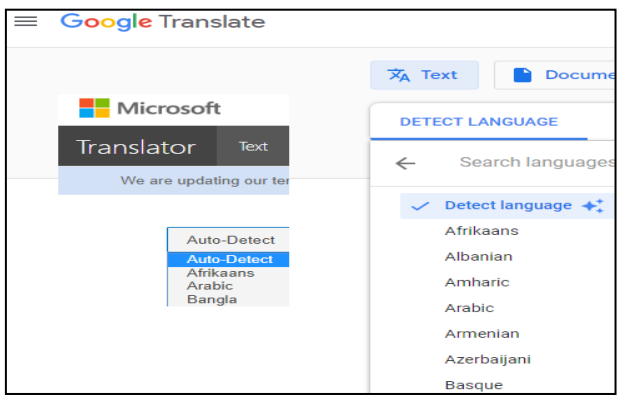


Fig. 2.Google and Bing Translator

for English to Assamese translation using n-gram model.

Phrase based systems consume high amount of memory and takes more time to translate. Hence, n-gram based approach has been embraced to overcome these constraints to some extent.

III. PROPOSED WORK

Most of the existing translation systems for English to Assamese and vice versa have used the phrase-based approach. But for the language pair with high dissimilarity in the formation of sentences, phrase based system may produce inefficient translation. Hence, n-gram based system has been proposed.

Both n-gram based and phrase-based systems use word-by-word alignment of the corpora. After alignment, the *tuples* or *translation units* are extracted for n-gram based system and for phrase based system the *phrases* are extracted for performing the translation. The following paragraph shows how the tuples and phrases are extracted from the aligned sentence.

From the Fig. 3 we can extract the tuples (t1…tn) as follows:

t1: Dispur_is#দিছপুৰ

t2: the_capital_of_Assam #অসমৰ্ৰাজধানী

Any pair of the source sentence words and the target sentence words satisfying the following two constraints can be termed as phrase.

- Words should be consecutive along both the sides of the bilingual phrase, and
- Words on either side of the phrase should not be aligned to a word which is out of the phrase [5], [8].

So, keeping these two constraints in mind, the extracted phrases (p1….pn) from Fig. 3 are as follows:

p1: Dispur_is#দিছপুৰ

p2: Dispur_is_the_capital_of_Assam#দিছপুৰ্অসমৰ্ৰাজধানী

p3: the_capital#ৰাজধানী

p4: the_capital_of_Assam#অসমৰ্ৰাজধানী

p5: of_Assam#অসমৰ

From the above, it is observed that the number of extracted tuples is significantly less than that of extracted phrases. It is clear that the sentence pair can be segmented into multiple phrases ([p1+p4], [p2]). But, only single segmentation is possible when extracting tuples ([t1+t2]). This makes phrase based approach more robust than the n-gram approach. But the problem is that, phrase based approach consumes more memory than the n-gram based approach as the number of phrase is more than that of tuple. Another problem is the search time. N-gram based system accomplishes a similar level of translation in less time than that of phrase based system.

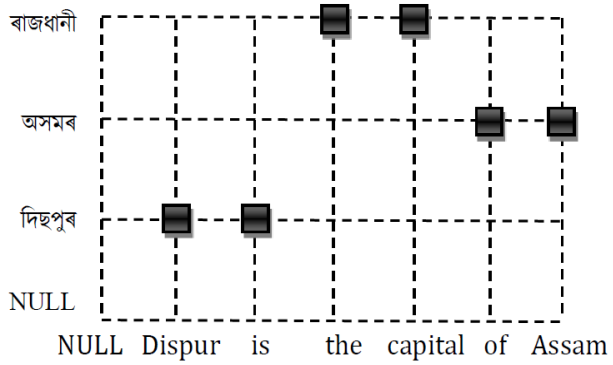


Fig. 3.Sentence alignment to show the extraction of tuple and phrase

A. *Background*

All statistical machine translation system essentially goes through three phases as follows:

- Language modeling: This phase involves the calculation of the probability of each word present in the target language corpus.
- Translation modeling: This phase takes care of the calculation of the probability of words present in the target language corpus given the probability of the words present in the source language corpus.
- Decoding: This phase involves the maximization of the probability to get the correct translation.

Phase-I: Language model

The language model provides the context of selecting the most prominent word to be placed after a sequence of words from many choices. The theoretical concept can be understood by this example. Consider an English sentence: *My phone is ringing*. One translated sentence in the Assamese language is মোৰ ফোন বাজি আছে In the translation process, after placing মোৰ, the next word to be placed may be ফোন or বাজি or আছে But during the probability calculation, probability of getting ফোন will be greater than বাজিand আছে i.e. P(ফোন) > P(বাজি) and also P(ফোন) > P(আছে). So the next word will be ফোন. After getting মোৰ ফোন, the probability of getting বাজি will be greater than that of আছে So the next word will be বাজি After মোৰ ফোন বাজি the probability of getting আছেwill be 1 (one), as there is no remaining word. i.e. P(আছে)=1. Therefore, আছেwill be put next to get the full translated sentence [12]. There is a provision of getting equal probability also. In that case any one of the equi-probable words can be put. For a single sentence of a particular language, there may have more than one translated sentence. In that case, word order plays an important role in handling the situation.

- **Word order**

Assume that in the above case, obtained translated sentences are মোৰ ফোন বাজিআছেমোৰ বাজিআছেফোন, মোৰ আছেফোন বাজিetc.. In this situation, there will be an ambiguity in producing the output. To handle this, the analysis of the text corpus is necessary as the appropriate order of the words to be placed in the sentences of a particular language can be learned from there. From the above three sentences, মোৰ ফোন বাজি আছেhas the appropriate order of each words compared to মোৰ বাজিআছেফোন and মোৰ আছে ফোন বাজি Hence, it is expected to get মোৰ ফোন বাজি আছে as the ultimate output [19]. Other than this problem, there is a chance to encounter another problem i.e. getting more than one meaning of a single word. In this case, word choice resolves the issue.

- **Word choice**

In the previously considered sentence, the word *phone* can have many synonymous words in Assamese like ফোন, দূৰভাষ দূৰালাপ দূৰভাষণী টেলিফোন etc.. Therefore, it is possible to get many translations based on these words. Since, each and every word in the text corpus has a specific probability; the choice of the word is done for the maximum probable word out of all the choices available on that corpus [19].

During the probability calculations in different steps of translation, *data sparsity* creates a major problem. Due to data sparsity, most possible word sequences may not be observed in training phase. As a result, the testing phase may show low efficiency. One solution is to make an assumption that the probability of a word depends only on the previous *n* words and this is called as n-gram model. The following section enlightens the n-gram model.

- **N-gram**

An *n*-gram can be termed as a contiguous sequence of *n* number of items from a particular sequence of text or speech [14]. Different applications may be having different types of items like phonemes, syllables, letters, words, base pairs etc.. An *n*-gram with size 1 is labeled as *unigram*, size 2 is labeled as *bigram* or *digram*, size 3 is labeled as *trigram* [14]. After these, the n-grams can be labeled as per the value of *n* like *four-gram*, *five-gram*, *six-gram* etc. [14], [9]. One example to understand this concept is given in Table- I.

○ ***N-gram models***

An *n*-gram model is based on the calculation of the probability and it is used to predict the next item for a sequence in the form of *Markov model* of order *(n - 1)*.

The n-gram models and the algorithms that use the n-grams have the following two main advantages:

- These are relatively simple and
- These are having the ability to scale up. It can be efficiently scaled up by simply increasing the value of *n* [12].

To make it clearer we can say that an n-gram model predicts Y_j by looking into $Y_{j-(n-1)}, \dots, Y_{j-1}$. Probability term for this is $P(Y_j | Y_{j-(n-1)}, \dots, Y_{j-1})$. Independent assumptions are made during the language modeling so that each word of a particular sentence is dependent only on the last $n-1$ words. These assumptions are important because:

- It significantly simplifies the problem associated with learning the language model from the data.
- The nature of the language is open and it is obvious to group the words together that are unknown to the language model.

It is to be noted that, in basic n-gram language model, the probability of a word from a sentence, with respect to some number of previous words from the same sentence (one word for bigram model, two words for trigram model, etc.) can be described as a *multinomial distribution* or *categorical distribution* [14].

Table- I: Demonstration of n-gram (up to 3-gram)

Type	Sample Sequence	1-gram sequence (unigram)	2-gram sequence (bigram or digram)	3-gram sequence (trigram)
Character	be_in_the_debt_of	b, e, _, i, n, _, t, h, e, _, d, e, b, t, _, o, f	be, e_, _i, in, n_, _t, th, he, e_, _d, de, eb, bt, t_, _o, of	be_, e_i, _in, in_, n_t, _th, the, he_, e_d, _de, deb, ebt, bt_, t_o, _of
Word	be in the debt of	be, in, the, debt, of	be in, in the, the debt, debt of	be in the, in the debt, the debt of

The probability $P(W_1, \dots, W_k)$, observing the sentence $W_1 \dots W_k$ is approximated in n-gram model as:

$$P(W_1, \dots, W_k) = \prod_{j=1}^{k} P(W_j | W_1, \dots, W_{j-1}) \approx \prod_{j=1}^{k} P(W_j | W_{1-(n-1)}, \dots, W_{j-1}) \quad (1)$$

where, W_k is the k^{th} word of the given sentence.

Using the n-gram frequency counts, the conditional probability can be calculated as follows:

$$P(W_j | W_{j-(n-1)}, ., W_{j-1}) = \frac{count\ (W_{j-(n-1)}, \dots, W_{j-1}, W_j)}{count\ (W_{j-(n-1)}, \dots, W_{j-1})} \quad (2)$$

This can be clarified with the help of the following example:

Let us consider a sentence *"I visited Gauhati University"*. The probability of this sentence using bigram $(n-2)$ can be approximated as-

P(I,visited,Gauhati,University)=P(I|<s>)P(visited|I)P(Gauhati|visited)P(University|Gauhati)P(</s>|University)

The approximation for the same sentence in trigram $(n=3)$ is-

P(I,visited,Gauhati,University)=P(I|<s>,<s>)P(visited|<s>,I)P(Gauhati|I,visited)P(University|visited,Gauhati)P(</s>|Gauhati,University)

It can be noted that, the first $(n-1)$ n-grams contexts are filled with $< s >$ known as start-of-sentence marker. Without the end-of-sentence marker $(</s>)$, the probability of a smaller sequence *I visited Gauhati* would always be higher than that of the longer sequence *I visited Gauhati University* [14].

○ ***Count smoothing***

During the counting of n-grams, we may encounter that some n-grams do not appear in the corpora. Those n-grams contribute 0 (zero) probability to the corpora. This may yield a value 0 (zero) while retrieving and if it is needed to multiply the probabilities with the n-grams having non-zero probabilities. To get rid of this, 1 (one) is added to the numerator and the number of types is added to the denominator. So, the new count will look like the following:

$$P(W_j | W_{j-(n-1)}, ., W_{j-1}) = \frac{count\ (W_{j-(n-1)}, \dots, W_{j-1}, W_j) + 1}{count\ (W_{j-(n-1)}, \dots, W_{j-1}) + t} \quad (3)$$

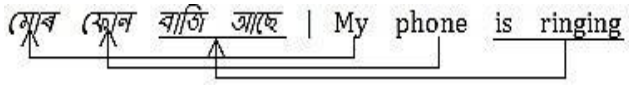


Fig. 4.Example of Sentence alignment

Where, t is the number of types present in the vocabulary. This makes sure that each n-gram has a minimum count of 1. Hence, a sequence that does not even occur in the corpora will also have a non-zero probability [19].

Phase-II: Translation model

Translation model aims to generate the sentence *d* of target language from the sentence *c* of source language by computing the conditional probability$P(d|c)$. It can be thought that the target sentence is being generated from source sentence word-by-word [12], [4]. The following example can be considered for better understanding:

If an English sentence *My phone is ringing* is translated to an Assamese sentence, one result is মোৰ ফোন বাজি আছে One possible alignment for the sentence pair can be as in Fig. 4. A number of alignments are possible for every sentence, such as word-by-word, phrase etc. For better understanding, the word-by-word alignment can be considered. If m numbers of words are present in source sentence and n numbers of words are present in target sentence, then $m \times n$ different alignments are possible. All connections towards the target position are equi-probable. Therefore, the order of words in the target sentence and the source sentence do not affect the probability of target given source. i.e. $P(target|source)$ [20].

After word-to-word alignment, the tuples are extracted. The bilingual units are generally referred to as the tuple. The probabilities of the translation model are approximated by the n-grams of the tuples at sentence level as follows:

$$P(T,S) \approx \prod_{m}^{M} P((t,s)_k \,|(t,s)_{k-1},(t,s)_{k-2},\dots\dots,(t,s)_{k-n+1}) \quad (4)$$

Where, t represents target and s represents source. The$(t,s)_k$ refers to k^{th} tuple of bilingual sentence pair [15].

Phase-III: Decoding

The process of probability maximization for the translated text can be termed as the decoding. The words having maximum likelihood are chosen. A target sentence search is performed that maximizes $P(T|S)$. i.e. $P(S,T) = arg\ max[P(S|T) \times P(T)]$ should be satisfied for getting the target sentence [12].

In the above formula, S and T denote source and target respectively. There is a possibility that a problem of infinite space search may arise in this case. To cope up with this problem, a method called *stacked search* is generally suggested. Here a partial alignment hypothesis list is maintained and the search is prompted to start with null hypothesis. That means we get the target sentence from an unknown source word sequence [12]. One example is as follows:

(অসমৰৰাজধানীদিছপুৰ| #), where, # is a place holder for an unknown source word sequence. As the search moves forward, one or more words are added to the hypothesis to extend the list of entries [12]. Example is as follows:

(অসমৰ ৰাজধানী দিছপুৰ| The Capital of Assam)

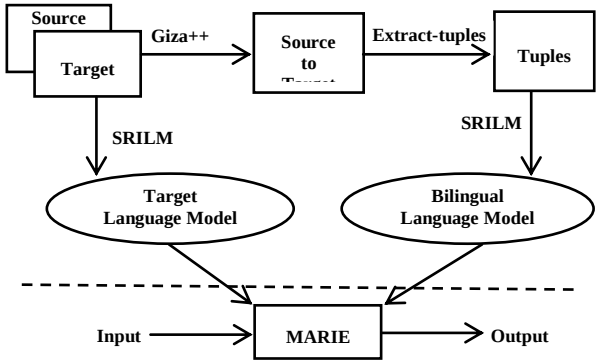


Fig. 5.Block diagram of the proposed system

The # can take any combination of the words in the list. But the search will terminate with a more promising complete alignment than any of the incomplete alignments [20].

B. *Architecture of the Proposed System*

The architecture of the proposed system (Fig. 5) has two parts. The upper part of the dotted line is for *training* and the lower part of the dotted line is for *testing* purpose. In the training phase the developed system is trained with the available data. In this case, the data is the English-Assamese parallel corpus. The more the data in the training phase, the better the efficiency of the system. Here, a corpus of about 15000 sentences has been used. From the architecture it can be said that the alignment of English words to Assamese words, the extraction of the tuples, creation of the target language model, creation of the bilingual language model and feeding these language models into the decoder falls under the training phase. The testing phase comprises the uses of 10 to 20 percentages of the training data to test the system, whether it is working properly or not. It is just an approximation that the system will work as desired. Basically, providing input to the decoder (MARIE) and getting output from the decoder falls under the testing phase.

From the block diagram (Fig. 5), it has been seen that parallel corpus of source and target language has been word-by-word aligned with the help of translation model toolkit called Giza++. From the aligned corpus the translation units are extracted. The translation units are called tuple in n-gram based statistical machine translation system. After extracting the tuples, the bilingual language model is created with the help of Stanford Research Institute Language Model (SRILM) toolkit. Also target language model is created with the help of SRILM from the target language corpus. The MARIE (n-gram based decoder) is then trained with the language models. After training, the decoder (MARIE) is ready to translate and accepts input from the user and yields the output accordingly. The details of the tools used in this work are described in the following section.

IV. EXPERIMENTAL SETUP

A. *Tools Used*

GIZA++ and SRILM have been used for translation model and language model respectively. For the decoding part MARIE has been used. The details regarding the tools have been given below:

GIZA++

In [16], it has been mentioned that GIZA was a part of the statistical machine translation toolkit called EGYPT. It was developed at the *Center for Language and Speech Processing* during the summer workshop in the year 1999 by the statistical machine translation team of Johns-Hopkins University [16]. GIZA was extended to GIZA++ by Fran Josef Och adding some new features [16]. The ideas about the extensions to GIZA which are included in GIZA++ can be found below:

- A new model named as *Model 4* [16].
- Another new model named as *Model 5* [16].
- Depending on word classes some Alignment models have been included [16].
- Implementation of the HMM alignment model: Baum-Welch training, Forward-Backward algorithm, empty word, dependency on word classes, transfer to fertility models etc. [16].
- New form for Model 3 as well as for Model 4 which allows the training of the parameter p_0 [16].
- For fertility, distortion or alignment parameters, many smoothing techniques have been included [16].
- Fertility models have been trained more efficiently than the previous ones [16].
- The pegging described in [4] has been accurately implemented. It is a series of heuristics to make the pegging efficient enough [16].

SRILM

It is a widely used language modeling toolkit. It calculates the n-grams of the corpus. The *n* may be of any value. Default value of n is 3. It requires huge monolingual corpus (In our case: Assamese) in well aligned manner. Also it can calculate the n-grams for bilingual corpus, such that the words are aligned and there is no gap between the aligned words.

The primary objective of SRILM is to support the *estimation* of the language model and *evaluation*. The creation of a model from the training data is termed as the estimation and the computation of the probability of a test corpus or corpora is termed as the evaluation.

The basis of SRILM is the n-gram statistics. Three main functionalities of SRILM are:

- To generate the count file of n-gram from the given corpus
- To train the language model with the help of the count file of n-gram
- To calculate the perplexity of the test data using the language model that has been trained.

For this system, SRILM has been used to get the target language model and also the bilingual language model. For target language model we have used Assamese text corpus.

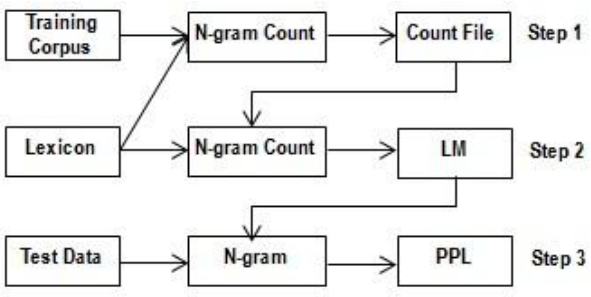


Fig. 6.Language model development steps

For bilingual language model we have used the file containing the translation units (tuple). These units are extracted from the alignment file of Giza++. After getting the tuples with the help of *extract-tuple* (a tool comes with MARIE), some modifications are needed to get the appropriate format for getting correct n-gram from the file. The SRILM tool calculates the n-gram in three steps as shown in the Fig. 6 [22].

MARIE

It is a statistical machine translation decoder based on the n-gram model. It implements the beam search strategy, which is based on the dynamic programming technique [15]. The decoding is guided by the source and it is performed monotonically [15]. While decoding, the partial translation hypotheses are generally arranged into distinct stacks as per the total number of source words they are covering [15]. The hypotheses covering the same source word competes with each other [15].

To tract the decoding at every step, MARIE uses *Threshold pruning* and *Histogram pruning* methods. In *Threshold pruning* method**,** each and every partial translation hypotheses having low score than a preset threshold value are eliminated [15], [7]. And in *Histogram pruning* method, the maximum number of partial translation hypotheses to be used is decided and the K-best ranked ones are considered [15]. *Hypothesis recombination* is also been allowed by MARIE. Here, exactly coinciding partial translation hypotheses in both the present tuple and the tuple trigram history are recombined [15].

B. *Steps Carried Out under the Proposed Work*

- **Creation and preprocessing of the corpus**

Parallel corpus of English and Assamese language is required for this translation system to work. Here, a parallel corpus of about 15000 sentences has been used. The domain of the corpus is mixed, comprising maximum part from tourism in India. The raw corpus may have some different format than the tool's format. That's why, before using the raw corpus, it is necessary to process it to get the desired format for the tools. If this is not done, the tools may not work properly and as a result it may not produce desired output. Some tools may need only the lowercase letters, some may need uppercase letters, some tools may require separate quotation mark etc.. Depending on the need of the tools, the data or the training data have to be processed first. Here, the following preprocessing steps are carried out:

- **Tokenization**

For correct alignment, the corpus should be in tokenized form. Tokenization means separation of the words, punctuations, extra spaces from the sentences of the corpus. This works by following some rule for the concerned language. The special rules are assigned in the *nonbreaking-prefix* file. For this purpose a perl script has been used. The perl script works only for some particular languages. So, for tokenizing Assamese corpus, some modification is needed. Since the structure of English and Assamese sentences is not same, hence direct use of the tokenizer results into break down of the composite words of Assamese sentence.

This is not desirable. Therefore, slight modification has been done into the code of the perl program to get the desired result for Assamese language. The commands for tokenization are:

./tokenizer.perl -l src<srccorpus>srccorpus.tok
./tokenizer.perl –l trg<trgcorpus>trgcorpus.tok

Here,
src→source language abbreviation (e.gen|as|hi|je|de etc.)
srccorpus → source corpus file along with the path
srccorpus.tok→ tokenized source file along with the path
trg → target language

abbreviation (e.gen|as|hi|je|deetc)
trgcorpus → target corpus file along with the path
trgcorpus.tok → tokenized target file along with the path

○ **Lower casing**

Lower casing of the words in the corpus is necessary for better result from Giza++. The uppercase letters are turned into lowercase letters using the inbuilt function of Ubuntu. The syntax for lower casing is:

tr '[:upper:]' '[:lower:]' <srcfile.tok>srcfile.tok.low

Here,
tr→ refers to translation, i.e translate from upper to lower case
[:xxxxxx:]→ this varies depending on the operation performed
srcfile.tok→ refers to tokenized file to be lowercased
srcfile.tok.low→ refers to output file after lowercasing.

Generally, this operation should be performed for both the source and the target file. But in Assamese language, there is no concept of lower and upper case. This language is not case sensitive. Yet this operation can be performed on the Assamese language corpus. This will not show any change in the corpus.

- **Creation of the translation model**

A major part of this system is the translation model. Basically, in this case, the calculation of the probability of the target sentence given the source sentence is performed. For this purpose Giza++ has been used. The details of creation of the translation model are discussed below:

○ **Translate the plain text to Giza format**

Giza++ works only on texts which are in Giza format. The plain text i.e. the tokenized and lowercased text is converted into Giza format with the help of the following:

./plain2snt.out srcfiletrgfile

Here,
srcfile→refers to the source language tokenized and lowercased file along with path
trgfile→refers to the target language tokenized and lowercased file along with path

This will create the files *srcfile.vcb*, *trgfile.vcb*, *srcfile_trgfile.snt*and*trgfile_srcfile.snt*. The descriptions of the files are as follows:

srcfile.vcb file contains the information from the source language corpus like, each word, frequency count and unique id.

trgfile.vcb file contains the information from the target language corpus like, each word, frequency count and unique id.

srcfile_trgfile.snt file contains each of the sentences from the parallel corpus (source and target) translated into the unique numbers for each of the words.

trgfile_srcfile.snt file contains each of the sentences from the parallel corpus (target and source) translated into the unique numbers for each of the words. It is just the opposite of *srcfile_trgfile.snt*and vice-versa is also correct.

○ **Create coocurrence file**

The coocurrence files are created from the above Giza format file. The syntax for creating the coocurrence file is as follows:

./snt2cooc.out srcfile.vcbtrgfile.vcbsrcfile_trgfile.snt >st.cooc

This will create the coocurrence file and write into the file *st.cooc*. The filename may be defined by the user i.e. it may be other than st.cooc.

○ **Make classes**

One package named as mkcls makes different classes of the text in the corpus based on similarity of the words. The number of class may be defined by the user. If the number of class is not defined, then the default number of class is created. The syntax of creating classes by mkcls is as follows:

./mkcls –C80 –n10 –psrcfile.tok.low –Vsrcfile.tok.low.vcb.classes
or
./mkcls –n10 –psrcfile.tok.low –Vsrcfile.tok.low.vcb.classes

The first command will create 80 numbers of classes using 10 numbers of rounds and the second command will create mkcls' default number of classes using 10 numbers of rounds. The number of rounds may vary. Higher the numbers of rounds better the result. This will create two files *srcfile.tok.low.vcb.classes* and *srcfile.tok.low.vcb.classes.cat* in the specified directory. The descriptions of the files created are as follows:

srcfile.tok.low.vcb.classes files hold the list of all words in alphabetical order along with the punctuations and also the frequency counts of all the words.

srcfile.tok.low.vcb.classes.cat files hold the list of frequencies and the collection of words for that corresponding frequency.

○ **Alignment of the corpus**

Once all the configuration files needed for the alignment are ready, the alignment process of the source and target corpus can be initiated. Since, .vcb files, .snt files, CoocurrenceFile and class files are obtained from earlier operations, Giza++ can align the corpus now. This is done using the following:

./GIZA++ -S srcfile.tok.low.vcb –T trgfile.tok.low.vcb –C srcfile_trgfile.snt –CoocurrenceFilesrc_trg.cooc

The above will align the source and target corpus with the help of source vcb, target vcb, source and target snt file and the coocurrence file. The files created by this are described below:

.a3.final file has a table with the below mentioned format:

$x\ y\ z\ t\ P(x|y,z,t)$

where,
x →Source sentence position
y →Target sentence position
z →Source sentence length

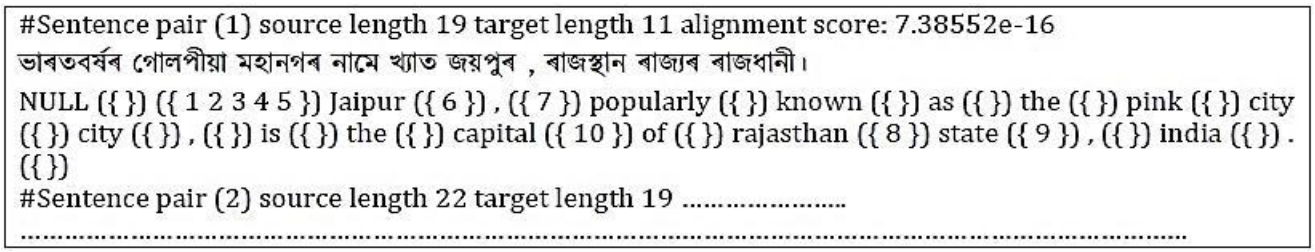


Fig. 7.Alignment Probability (.A3.final) file

t →Target sentence length

$P(x|y,z,t)$ →The

probability that a source word in the position x is moved to the position y in a sentence pair with length $\boldsymbol{z}$ and $\boldsymbol{t}$[18]

.A3.final contains-

The matches for the source sentence to the target sentences and give the match an alignment score. It is the desired alignment file. It creates three lines for a particular sentence. It shows the number of words present in the source sentence and also in the target sentence along with the alignment score (Fig. 7).

Source corpus words, their unique id and frequency counts are shown in *.trn.src.vcb*. It matches with *srcfile.vcb*.

Target corpus words, their unique id and frequency counts are shown in *.trn.trg.vcb*. It matches with *trgfile.vcb*.

*.tst.src.vcb*is empty.

*.tst.trg.vcb*is empty.

.perp file contains the list of perplexity for each iteration and model. The form of the file is as follows:

"trsz tssz itr mdl tr_pp ts_pp tr_vit_pp ts_vit_p"

where,

- *trsz*→size of training data
- *tssz*→ size of test data
- *itr*→ no. of iteration
- *mdl*→model
- *tr_pp*→perplexity of the training
- *ts_pp*→perplexity of the test
- *tr_vit_pp*→viterbi perplexity of the training
- *ts_vit_pp*→viterbi perplexity of the test

.d3.final is alike *.a3.final* but the position of x and y are switched.

.n3.final is having the probability of all the source tokens with 0 fertility, 1 fertility,……., N fertility.

The table created after all the iterations of Model 4 training is kept in *.t3.final*.

The probability for the insertion of null after a source word is kept in *.p_03.final*.

The translation table associated with Model 4 is kept in *.d4.final*.

The distortion table associated with IBM-4 is kept in *.D4.final*.

The parameter settings used in the training are stored in *.gizacfg*.

○ **Extraction of the translation units (tuples)**

For n-gram based translation system, translation units (often called as tuple) are extracted. A tool called as *extract-tuples* comes with MARIE. This is used to extract tuples from the aligned file. i.e from the *.A3.final* file. But this tool cannot provide the tuples in correct format as the SRILM tool needs to create the bilingual language model. This extracts the tuples as follows [6]:

srcword # trgword # 0

where,

srcword→ denotes the source word

trgword→ denotes the target word

The # between the words experiences two spaces before and after. Also an extra # 0 comes with it. This is not desirable. If we apply SRILM to count the n-gram on this type of file, then SRILM will break this into srcword, <space>, #, <space>, trgword, <space>, #, <space>, 0. And will calculate the n-gram for all of these. To get rid of this, some modification in the file is needed. The "<space>#<space>0" has been removed and also the "<space>#<space>" has been replaced by "#". Then n-gram has been calculated for the file (Fig. 8).

- **Language model creation**

After getting the translation model, creation of language model is necessary for training the system. Without the language model, working of statistical machine translation system is not possible. Creation of language model means calculation of the n-gram of each word in the corpus. For this system a tool called SRILM has been used to calculate the n-gram of the words in the corpus. The n-gram has to be calculated using the following:

./ngram-count –order n –text file1 –lm file2

or

./ngram-count –text file1 –lm file2

The first one calculates the n-gram of order n. The value of n is user defined. The second one does not have the order field. In this case SRILM will calculate up to 3-gram. 3-gram is the default value of SRILM. *file1* is the input file and the *file2* is the output file, i.e the language model file. The format of the output file is ARPA (Advanced Research Projects Agency) [22].

Two types of language model have been created in this case. These are: *Target language model* and *Bilingual language model*. These models are discussed below.

○ **Target language model**

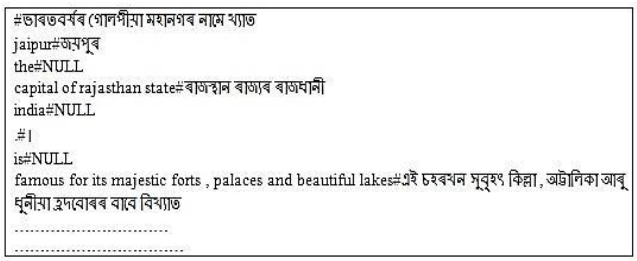


Fig. 8.Extracted tuples after modification

Target Language model is the language model created using the target language corpus. In this case it is Assamese.

The corpus is first tokenized and lowercased then target language model is created using SRILM. As it is mentioned previously that Assamese is not a case sensitive language, therefore, lowercasing this corpus can be skipped.

○ **Bilingual language model**

Bilingual Language model has been created using the file containing the tuples extracted from the aligned corpus. The parallel corpus of English and Assamese has been processed using Giza++ to get the desired aligned file. The aligned file is then processed through the extract-tuples and the tuples are extracted. Here, the extracted tuples were in the decoder's format. Therefore, one more step mentioned under the heading *Extraction of the translation units (tuples)* has been implemented to rectify this problem. After that the bilingual language model has been created.

- **Decoding**

After getting the language models, the decoder called MARIE has been trained. Once the decoder has been trained the system is ready to translate. The training of the decoder is very easy. Simply copy the language models to the directory of MARIE. Then take the test file of source sentences to be used as the input file for the system. The output is obtained by the following:

./marie-v1.1 –i inputfile –fBMbilingualLMfile

or

./marie-v1.1 –i inputfile –fBMbilingualLMfile –fTMtargetLMfile

The first one does not contain the target language model file. It yields the output based on the Bilingual Language model only. The words in the input file are matched by searching the Bilingual Language model file. The second one contains the Bilingual Language model file and the Target Language model file. The use of target language model file can help in producing better result. Decoding part is mainly to produce the result. The produced result is discussed in the *results and discussions* section.

V. RESULTS AND DISCUSSIONS

The system was trained with almost 15000 sentences from parallel corpus of English and Assamese. In the testing phase, 15% of the training data have been used. Table - II shows some of the translated sentences.

Some sentences are correctly translated and some of them are partially translated. The system has been tested using four orders in the SRILM. The order means the level up to which the n-gram is being calculated during the language model creation by SRILM. The output was produced by using only Bilingual Language Model (BLM) and also by Bilingual Language Model and Target Language Model (TLM) together. For all those produced output, BLEU score has been calculated and the variation is shown in the Fig. 9. It has been seen that, for orders 1, 2 and 3, the system performs same for only BLM and also for BLM & TLM together. But for the order 4, translation using only BLM is performing better than that of the translation using BLM and TLM together.

The system is able to translate the English text to Assamese text, but the translated Assamese text is not up to the mark of desired translation. The problems we have noticed in the translated text are as follows:

- Some words of the input sentence are not translated at all
- Some proper nouns are not translated

The first problem arises because of the alignment of some words to NULL. It is due to the high dissimilarity in the formation of the sentences and smaller corpus. Secondly, some of the proper nouns are translated as desired but some

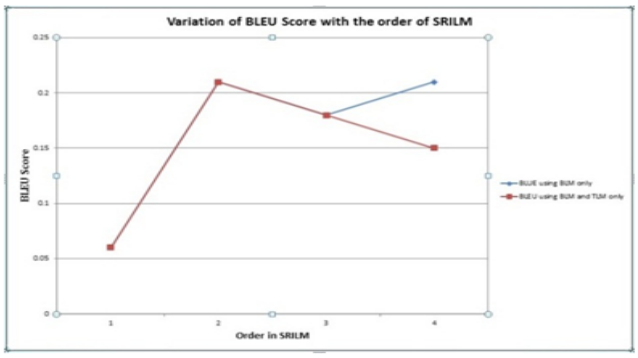


Fig. 9.Variation of BLEU score with the order of SRILM

are left out. One possible solution to the left out proper nouns is to translate the letters directly like A→অ, B→ব, C→চetc.

By increasing the training data set, it can be expected that the system will produce better result than the current result.

Table- II: Output from the user defined input

Input Sentence	Output Sentence
Mizoram is a land of great natural beauty.	মিজোৰাম প্ৰা কৃতিক সৌন্দৰ্যৰ ৰাজ্য।
Khimsar is a fine tourist destination in Rajasthan.	Khimsar ৰাজস্থানৰ এখন সুন্দৰ পৰ্যটনস্থলী।
Mumbai is a cluster of seven islands.	মুম্বাই সাতটা দ্বীপৰ সমষ্টি।

VI. CONCLUSION

The demand for the machine translation is increasing day-by-day. The machine translation for English-Assamese language pair is lagging behind the other language pairs. Assamese speaking peoples of northeastern India are not able to access many facilities due to language incomprehensibility. Keeping this in mind, a noble approach to translate English text to Assamese text using n-gram model has been initiated. Though this system has also been developed using phrase based decoder, yet the same system has been developed using different decoder hoping to get a better system in terms of the speed and less memory consumption. The performance and quality of translation vary with the change of used method and the dataset. In recent research, the accuracy of translation has shown significant improvement with the use of neural network and deep learning techniques.

ACKNOWLEDGMENT

The English-Assamese parallel corpus has been provided by the department of Information Technology, Gauhati University, Gopinath Bordoloi Nagar, Jalukbari, Guwahati, Assam, India.

REFERENCES

1. M. V. M. Barkade, P. R. Devale, "English to Sanskrit machine translator (lexical parser)", International Journal on Computer Science and Engineering, vol. 02, no. 06, pp. 2084-2091, 2010.
2. A. K. Barman, J. Sarmah, S. K. Sarma, "Assamese WordNet based Quality Enhancement of Bilingual Machine Translation System", in Proceedings of the Seventh Global Wordnet Conference, University of Tartu press, Jan. 2014, pp. 256-261.
3. K. K. Baruah, P. Das, A. Hannan, and S. K. Sarma, "Assamese-English bilingual machine translation", International Journal on Natural Language Computing, Available: https://doi.org/10.5121/ijnlc.2014.3307.
4. P. F. Brown, S. A. D. Pietra, V. J. D. Pietra, R. L. Mercer, "The mathematics of statistical machine translation: parameter estimation", Association for Computational Linguistics, vol. 19, no. 2, pp. 263-311, 1993.
5. J. M. C. Crego, J. B. M. Acebal, "Architecture and modelling for n-gram-based statistical machine translation", Ph.D dissertation, TALP Research Center, Speech Processing Group, Department of Signal Theory and Communications, Universitat Polit`ecnica de Catalunya, Barcelona, 2008.
6. J. M. Crego, M. R. Costa-jussa, J. B. Marino, and J. A. R. Fonollosa, "Ngram-based versus phrase-based statistical machine translation", TALP Research Center, Universitat Polit`ecnica de Catalunya, Barcelona, 2014.
7. J. M. Crego, J. B. Marino, and A. de. Gispert, "An ngram-based statistical machine translation decoder". TALP Research Center, Universitat Polit`ecnica de Catalunya, Barcelona, INTERSPEECH, 2005.
8. J. M. Crego, F. Yvon, "Factored bilingual n-gram language models for statistical machine translation", Mach Translat, Available: https://doi.org/10.1007/s10590-010-9082-5.
9. P. Das, K. K. Baruah, "Assamese to English statistical machine translation integrated with a transliteration module", International Journal of Computer Applications, vol. 100, no. 5, pp. 0975-8887, 2014.
10. P. Das, K. K. Baruah, A. Hannan, S. K. Sarma, "Rule based machine translation for Assamese-English using apertium", International Journal of Emerging Technologies in Computational and Applied Sciences, vol. 8, no. 5, pp. 401-406, 2014.
11. G. V. Garje, G. K. Kharate, "Survey of Machine Translation Systems in India", International Journal on Natural language Computing, Available: https://doi.org/10.5121/ijnlc.2013.2504.
12. A. Hannan, S. K. Sarma, Z. Hussain, "Marie: A statistical approach to build a machine translation system for English Assamese language pair", International Journal of Computer Sciences and Engineering, Available: https://doi.org/10.26438/ijcse/v7i3.774779.
13. J. Hutchins, "Multiple uses of machine translation and computerised translation tools", International Symposium on Data and Sense Mining, Machine Translation and Controlled Languages, 2009.
14. D. Jurafsky, J. H. Martin, "N-gram language models". In (Third Edition draft), Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, pp. 37-62, 2018.
15. J. B. Marino, R. E. Banchs, J. M. Crego, A. de. Gispert, P. Lambert, J. A. R. Fonollosa, M. R. Costa-jussa, "N-gram-based machine translation", Association for Computational Linguistics, vol. 32, no. 4, pp. 527-549, 2006.
16. F. J. Och, H. Ney, "A systematic comparison of various statistical alignment models", Computational Linguistics, vol. 29, no. 1, pp. 19-51, 2003.
17. M. D. Okpor, "Machine translation approaches: Issues and challenges", International Journal of Computer Science, vol. 11, no. 5, pp. 159-165, 2014.
18. M. A. Sati, "Word alignment using Giza++ and Cygwin on windows", International Journal of Engineering Research & Technology, vol. 2, no. 5, pp. 1762-1765, 2013.
19. M. T. Singh, R. Borgohain, S. Gohain, "An English-Assamese machine translation system", International Journal of Computer Applications, vol. 93, no. 4, pp. 1-6, 2014.
20. N. Sharma, P. Bhatia, V. Singh, "English to Hindi statistical machine translation system", ME thesis, Thapar University, Patiala, 2011.
21. J. Slocum, "A survey of machine translation: Its history, current status, and future prospects", Computational Linguistics, vol. 11, no. 1, pp. 1-17, 1985.
22. A. Stolcke, "SRILM-An extensible language modeling toolkit", Speech Technology and Research Laboratory, SRI International, Menlo Park, CA, U.S.A., 2004.
23. P. Unnikrishnan, P. J. Antony, Dr. K. P. Soman,"A novel approach for English to South Dravidian language statistical machine translation system", International Journal on Computer Science and Engineering, vol. 02, no. 08, pp. 2749-2759, 2010.

AUTHORS PROFILE

Zakir Hussain is a research scholar in the department of Computer Science and Engineering at National Institute of Technology (NIT) Silchar. He has completed his M.Tech. in Information Technology from Gauhati University in the year 2015. He has completed his BE (Honours) in Computer Science and Engineering in the year 2012 and bagged 4th position in the university. He was the state topper in the subject of General Science in board examination conducted by Assam state board in the year 2005.He was a Guest/Part time faculty in the level of Assistant Professor at Barak Valley Engineering College; a government of Assam institution in the district of Karimganj, Assam. He has 4+ years of teaching and administrative experiences in government run educational institutions and offices. His research interest is in the field of Natural Language Processing, Machine Learning, Data Analytics, and Data Mining.

Dr. Malaya Dutta Borah is working as an Assistant Professor in the Department of Computer Science & Engineering at National Institute of Technology (NIT) Silchar, Assam. She has completed her PhD in Computer Science and Engineering from Delhi Technological University (Formerly Delhi College of Engineering), Delhi. She has completed her M.E. (with distinction) in Computer Technology and Applications from Delhi College of Engineering, Delhi and B.Tech. in Computer Science and Engineering from North Eastern Regional Institute of Science and Technology, Arunachal Pradesh. Before joining NIT, she worked at Assam Engineering College, Delhi Technological University, Inderprastha Engineering College. She has authored/co-authored around 30 research papers in national/international journals/Conferences. She is actively involved in research works in the field of Data Mining, Cloud Computing, ICT and e-governance. She is the associate member of CSI (India) and IEEE. For more information, kindly visit: http://cs.nits.ac.in/malaya/

Abdul Hannan has completed his MSc. in Computer Science from Gauhati University in the year 2009 and M.Tech. in Information Technology from Gauhati University in the year 2012. He is currently working as a faculty in Gauhati University. He is pursuing his PhD in Machine Translation for Assamese Language. He has 7+ years of teaching experience in various educational institutions. His area of Interest is NLP, Computer Networks, and Machine Learning.

View publication stats

Journal of The Gujarat Research Society

ISSN: 0374-8588

Volume 21 Issue 15, December 2019

Optimization of FSW Process parameters for achieving a goodquality joint

Satyaveer Singh, N. Yuvraj, Reeta Wattal

Department of Mechanical Engineering, Delhi Technological University, Delhi, India

Abstract

Friction stir welding is basically a solid state joining technique that is widely used for joining Aluminum alloys in aerospace, automotive and many other applications of commercial importance. In the present study, dissimilar materials plates were FS welded by varying only pin offset and other parameters such as rotational speed, weld speed and tool tilt angle and plunge depth are made to be constant. The mechanical properties (Hardness and Tensile strength) of the Dissimilar Friction Stir welded (DFSW) specimens were tested and compared with the base materials. The observations have been elaborated in detail and it is observed that the weld parameters have a significant effect on mechanical and microstructural properties of the welds.A cylindrical pin profile is adopted as its geometry had been proven to yield better weld strengths.

Keywords: AA 8011 and 99.65% Cu, Friction stir welding (FSW)

1. Introduction

Dissimilar materials joints are widely used in power generation, chemical, petrochemical, nuclear, aerospace, transportation, and electronics industriesdue to their technical and beneficial advantages[1]. Basically Aluminium(Al) and copper (Cu) are the two common metals in the electrical power industry. Al–Cu transition pieces are widely used totransmit theelectricity.Due to differentchemical, mechanical and thermal properties of materials, dissimilar materials joining presentmore challenging problems than similar materials joining by friction stir welding (FSW).However, the dissimilar combination of Al and Cu is generally difficult for fusion welding. Therefore, the solid-state joining methods, such as friction stir welding, roll welding, and explosive welding have received much attention. The relative motion between the tool and the substrate generates frictional heat that creates a plasticized region around the immersed

Journal of The Gujarat Research Society
ISSN: 0374-8588
Volume 21 Issue 15, December 2019

portion of the tool. The tool shoulder prevents the plasticized material from being expelled from the weld. The tool is traversed along the joint line, forcing the plasticized material to coalesce behind the tool to form a solid-phase joint.The microstructural evolutions after the FSW process are characterized by three zones: the Stir Zone (SZ), the Thermo-Mechanically Affected Zone (TMAZ), and the Heat-Affected Zone (HAZ).

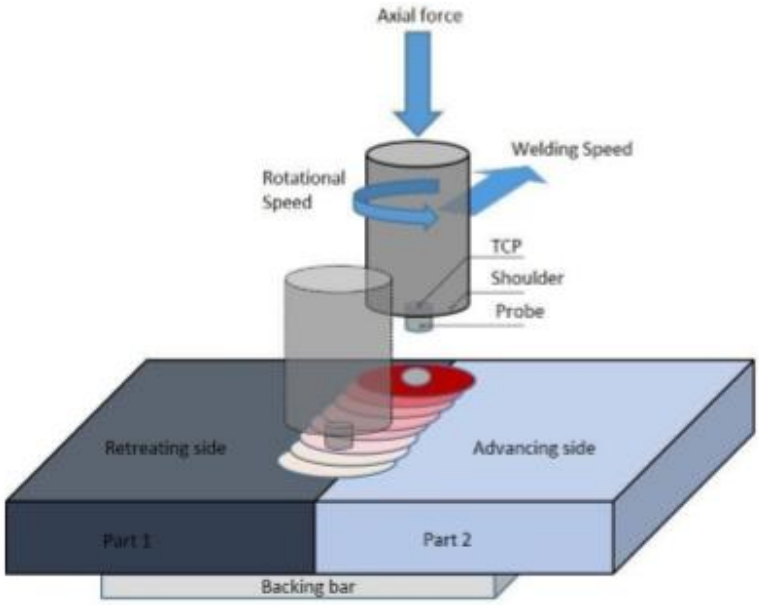


Fig.1. Schematic of Friction Stir Welding

It was reported that sound dissimilar FSW Al–Cu joints were difficult to achieve, and the joints usually failed at the nugget zone or along the interface between the twomaterials during the mechanical tests. The poor weld formed is due to the ability of the various brittle IMCs formed in the nugget zone [2-4, 20]. The formation of Al_4Cu_9 and Al_2Cu could be formed at the joint interface [5]. A sound FSW Al–Cu joint could be obtained by offsetting the tool to the Al side under a lower heat input condition [6, 18]. It is well documented that many parameters, such as tool offsetting, rotation rate and traverse speed, influenced the weld properties of the dissimilar FSW joints [7-8, 19]**.** Limited studies have been performed on the FSW dissimilar Al–Cu joints till now, and systematic research on the FSW parameters is still lacking. Therefore, it is worthwhile to study the effect of FSW process parameters on the microstructure and mechanical properties of the dissimilar FSW Al–Cu joints. In this paper, various FSW butt Al–Cujoints were

produced under different pin offsets and the microstructure and mechanical propertieswere studied in detail. The purpose of this study is to find out the feasible combination of FSW process parameters to perform defect free welding of dissimilar materials and to investigate the effect of the tool offset as a strategy to weld dissimilar materials.

2. Experimental procedures

8011Aluminum and commercially pure copper (99.65% purity) plates 3mm in thickness, 180mm in length, and 45mm in widthwere butt-welded using a heavy-duty vertical milling machine adapted for FSW: Bharat Fritz Werner (BFW) with spindle motor capacity 11KW asshown in the fig. 2.The welding tool used in this study was made of tungsten carbide and had a shoulder 20mmin diameter and a pin 6mmin diameter and 2.65mm in length as shown in the fig. 3.And thechemical composition is given in Table 1.Welds were made with a clock-wisely rotating pin at arotation rate of 750rpm and a constant traverse speed of 100 mm/min.Tool tilt angle of 2 degrees was used in the present study.The process parameters of the FSW taken up for the experiment and their levels are given in the (Table2).

Table 1

Shows the Chemical Composition of the metals

Material	Al	Cu	Mg	Si	Fe	Ni	Mn	Zn	Sn	Pb	Ti	Cr	V
Al 8011	98.50	0.103	0.086	0.231	0.710	0.012	00.132	0.160	0.004	0.019	0.012	0.021	<0.01
Copper	----99.65-	- -	- - -	0.001	0.235	----	0.012	0.037	---	----	----	----	

Table - 2

Factors and their Levels

Parameters	Levels			units
	Level 1	Level 2	Level 3	
Tool Pin offset	0.5	1.0	1.5	mm
Rotational speed	550	750	950	rpm

During the welding processes, several pin offsets from 0mm to 5mm were used. For convenience, in this study, an offset of 0mm denotes the positionwhere the pin just located at the butt line. Microstructural characterization and analysis were carried out usingOptical microscope:*Mitutoyo,Japan*. Tensile tests of 3mm thick specimens having 12mm gauge length were performed according to ASTM E8M standard at room temperature at a crosshead speed of 3mm/min. The tensile test specimens were prepared by Wire EDM and tested on the Tensometer.

Fig.2. Experimental setup of FSW **Fig.3.**Tool used in present study

To perform Friction stir welding and testing of dissimilar material welded samples, following machines/equipments were required; Shaper machine for dimensioning the work piece, A heavy-duty vertical milling machine adapted for FSW: *Bharat Fritz Werner (BFW)*,FSW fixture, Wire EDM for making sample for testing : *Steer Corporation,China*, Tensometer for Tensile testing: *Kudale Instruments Pvt. Ltd.,Pune*, Micro-hardness testing machines: *Mitutoyo, Japan*, Metallurgical polishing machine for grinding and polishing,Optical microscopes for microstructure: *Mitutoyo,Japan*, Steriozoom microscope for macrograph: *Mitutoyo,Japan.*

Journal of The Gujarat Research Society
ISSN: 0374-8588
Volume 21 Issue 15, December 2019

3. Results and discussion

3.1. The effect of fixed location on the weldability of the joints

Whenever the stronger material and the soft material were friction stir welded together, previous studies indicate that the weld quality was clearly influenced by the fixed location [9].For a sound weld to occur the stronger material should be fixed at the advancing side and the softer material should be placed at the retreating side. In this case the Cu plate is the stronger material and so it was placed on the advancing side.

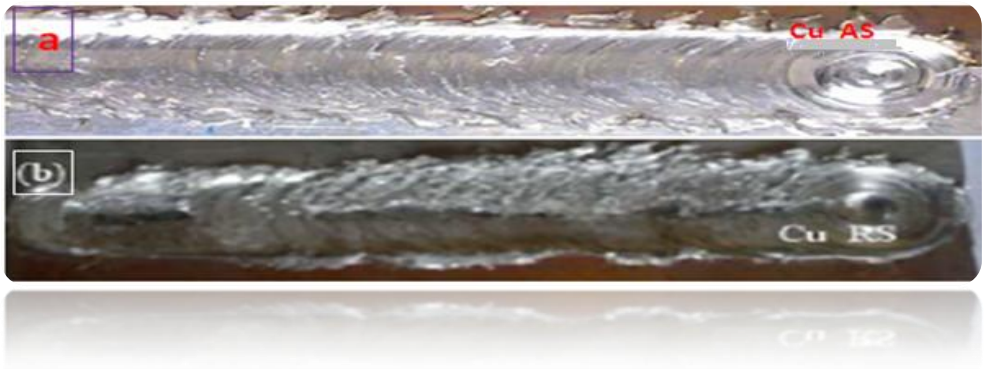


Fig.4. Surface Morphologies of FSW weld joints showing (a) Cu at the advancing side(b) Cu at the retreating side.

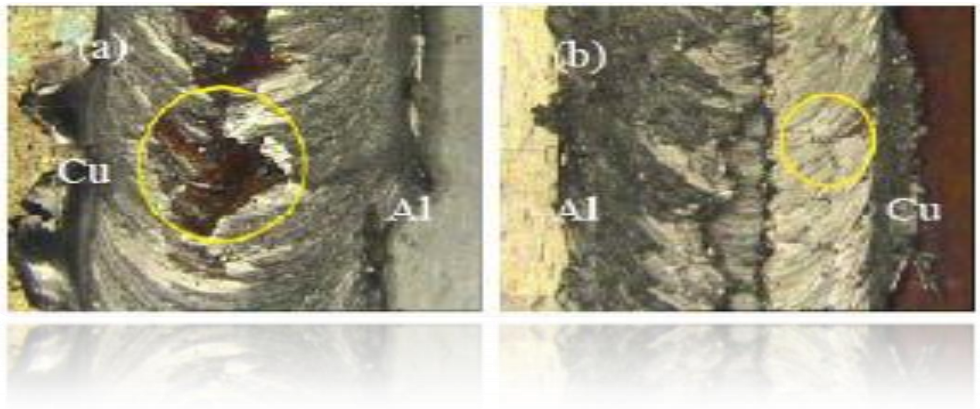


Fig.5. Surface morphologies of FSW weld jointsshowing (a) Macro Cracks
(b) Micro Cracks [15]

This study shows that even though cracks were observed in both the positions of the Cu plate, a

better mixing of the Al-Cu was found when Cu was kept at the advancing side. When the Cu plate was fixed at the retreating side, the hard material not seems to have reacted in the welding process. When the Cu plate was fixed on the retreating side certain macro cracks were observed but when the Cu plate was fixed on the advancing side certain micro cracks have been observed as shown in the fig.5. Surface morphology of the FSW Al–Cu joints for the different fixed locations at a welding parameter of 750 rpm – 100mm/sec is shown in the fig.4. It is clear that when the Cu plate was fixed at the advancing side, sound weld surface was achieved, as shown in Fig.4(a). However, when the Cu plate was fixed at the retreating side, the weld surface was very poor and certain macro cracks were observed.In a study of FSW Al–Fe joints, Watanabe et al. [9]reported that a long crack line was observed on the weld surface when the pin rotation direction was counter clock-wise, and sound weld surface could be obtained in a reversed rotating direction.

3.2 The effect of pin offset on the weldability of thejoints

In FSW of dissimilar materials, the pin offset is an important parameter influencing the weld quality [10, 11, 12, and 13]. The results in this study and most previous studies [10, 14]proved that sound dissimilar welds would be produced under a larger pin offset to the softer material. This should be related to the huge differences in physical, chemical and mechanical properties between the dissimilar materials. When the pin offset was larger towards the softer material, only a few Cu pieces with relatively small size were moved from the Cu bulk. It was easy for the small Cu piece to mix into the Al base and react with the Al base in the nugget zone, and therefore sound metallurgical bonding would be obtained at the Al–Cu interface. On the other hand, many large Cu pieces were stirred into the nugget zone at a smaller pin offset. The Cu pieces were harder than the Al matrix, therefore, the large Cu pieces were hard to deform and flow in the Al matrix, and the mixing between the large Cu pieces and the Al matrix would be very difficult. This led to the poor surface bonding and the formation of many voids has been shown in Fig.6. Moreover, when the pin offset was smaller, more Al–Cu IMCs have been formed because the more Cu pieces were stirred into the nugget zone. Thus, the joining between the Al and Cu became poor due to the brittle nature of the IMCs.

Journal of The Gujarat Research Society
ISSN: 0374-8588
Volume 21 Issue 15, December 2019

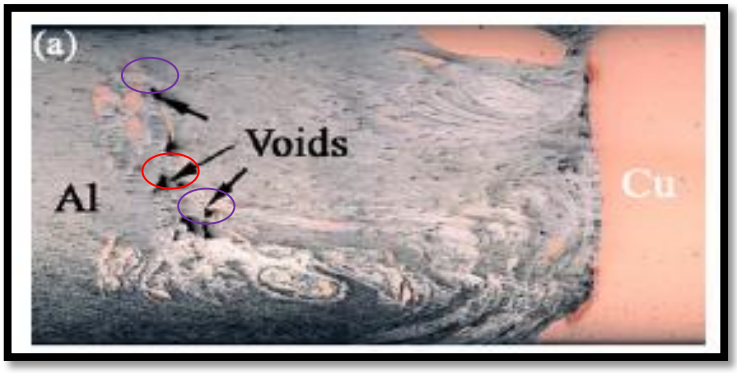


Fig.6. Cross-sectional macrostructure of the joints at rotation rates of (a) 400 rpm. [15]

During the FSW process, the materials were transported from the retreating side to the advancing side behind the pin where the weld was formed [16]. The hardness of the Cu is larger than that of the Al, and the pin stirred mainly in the Al base metal during FSW, so the material flow occurred mainly in the soft Al base metal. If the stronger material was fixed at the retreating side, the hard material was difficult to transfer to the advancing side because the hard material hardly flew. In this case, a large volume defect would form and the excessive soft material would be extruded out from the nugget zone. However, when the softer material was fixed at the retreating side, the soft material was transported to the advancing side easily, and the material flow cycle in the nugget zone was performed normally.

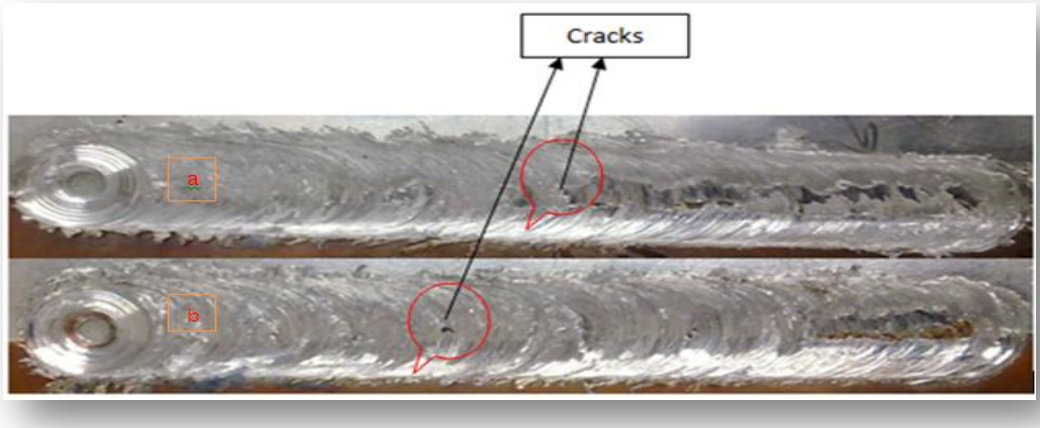


Fig.7. Surface morphologies of the FSW Al–Cu joints under a welding parameter of 750 rpm-100mm/min for pin offsets of (a) 0mm, (b) 0.5mm

In the present research the surface quality became poorer under the smaller pin offsets. Many cracks were observed obviously on the weld surface for a pin offset of 0mm and 0.5mm as shown in Fig.7.Sound defect-free joints could be obtained at a larger pin offset of 1.0 mm when pin is offsetting towards Al side and the nugget zone consisted of a mixture of the Al matrix and Cu pieces. Therefore, sound defect-free welds could be achieved only at larger offsets, no less than 1mm in this study.

3.3The effect of rotational speed on the weldability of the joints

Fig.8.shows the surface morphologies of FSW Al-Cu joints under different rotation rates. As shown fewer cracks were seen under a lower rotation rate of 600 rpm but at 950 rpm and 1200 rpm sound weld was not achieved. This might be due to the formation of inter-metallic compounds under the enhanced reaction between Al and Cu. Under the rotation speed of 1200 rpm many macro cracks were observed. When the rotation of the tool pin was high, large pieces ofCuwould be detached from the bulk and get distributed in the bottom and the retreating side of the nugget zone. When the rotation speed was fixed to 750 rpm small pieces of Cu would be scratched off from the bulk and thus at a certain portion of the weld zone seem to have mixed properly. The increase in rotational speed from 750 to 1200 rpm resulted in a harsher material flow. In Fig.8 (d) the distribution of particles along the onion rings in composite structure does not exist anymore. Also the majority of particles will increase than that seen at 750 rpm.

Journal of The Gujarat Research Society
ISSN: 0374-8588
Volume 21 Issue 15, December 2019

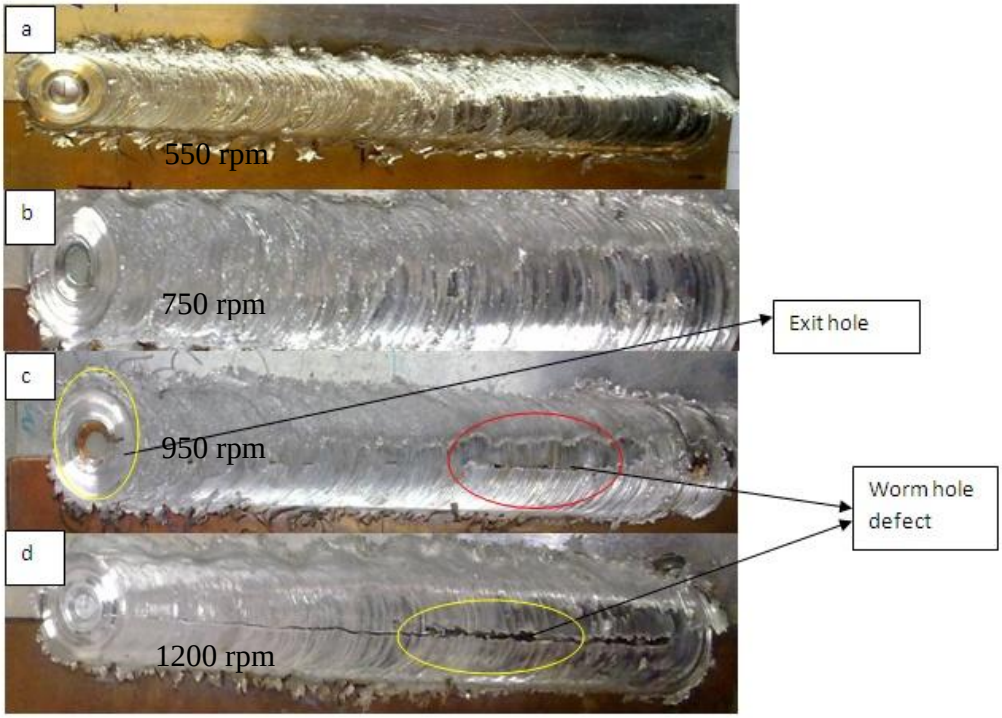


Fig.8. Surface morphologies of the FSW Al–Cu joints under a welding parameter of 100 mm/min and rotational speed of (a) 550 rpm (b) 750 rpm (c) 950 rpm(d) 1200 rpm

3.4 Tensile testingobservation

Tensile tests of 3mm thick specimens having 12mm gauge length were performed according to ASTM E8M standard at room temperature at a crosshead speed of 3mm/min in order to determine the mechanical properties (ultimate tensile strength and Elongation) of the welded and base materials. Thereduced section length is 66 mm and its width is 10mm. The tensile test specimens were prepared by Wire Electro Discharge Machining and tested on the Tensometer.Properties that are directly measured via tensile test are ultimate tensile strength, maximum elongation and reduction in area. Fig. 9 shows tensile testing specimen.Mechanical properties of the base metal were shown in table 3.

Journal of The Gujarat Research Society
ISSN: 0374-8588
Volume 21 Issue 15, December 2019

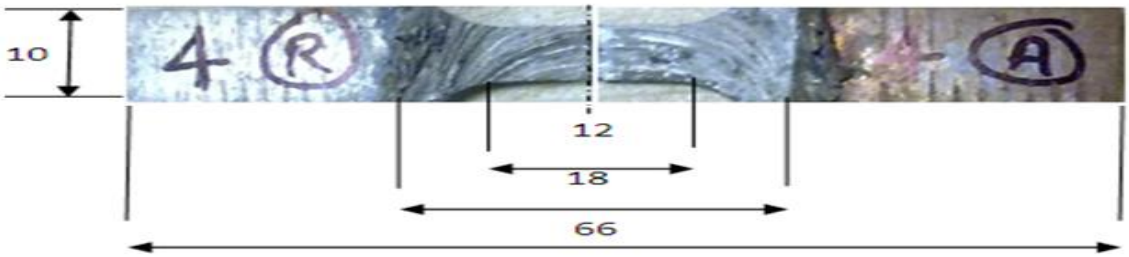


Fig.9. TensileTesting specimen

Table 3

Mechanical properties of the base metal

Base metal	Tensile strength (MPa)	Elongation (%)
Al 8011	127.3	8.56
Cu pure	350.4	5.67

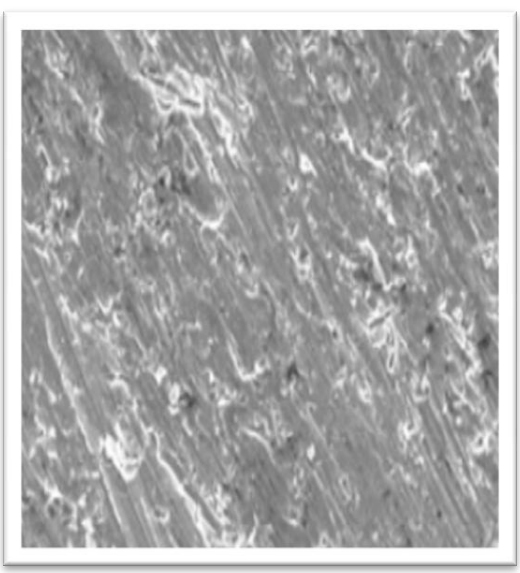


Fig. 10. Microstructure of AA8011-H18**Fig. 11.**Microstructure of pure copper

Microstructure of Al 8011 and copper is shown in the fig.10 & fig.11 respectively.

Journal of The Gujarat Research Society
ISSN: 0374-8588
Volume 21 Issue 15, December 2019

Table4

Mechanical properties and fracture locations of the welded jointsin transverse direction to the weld center line.

Pin offset (mm)	Tensile properties at room temperature		
	Tensile strength (MPa)	Elongation (%)	Fracture location
Cu side			
1.5	43.6	2.75	TMAZ of Cu
1.0	114.6	3.25	TMAZ of Cu
0.5	136.6	2.50	Nugget
0	105.0	4.16	Nugget
0.5	139.0	6.63	TMAZ of Al8011
1.0	145.0	4.00	TMAZ of Al8011
1.5	134.0	1.50	TMAZ of Cu
Al side			

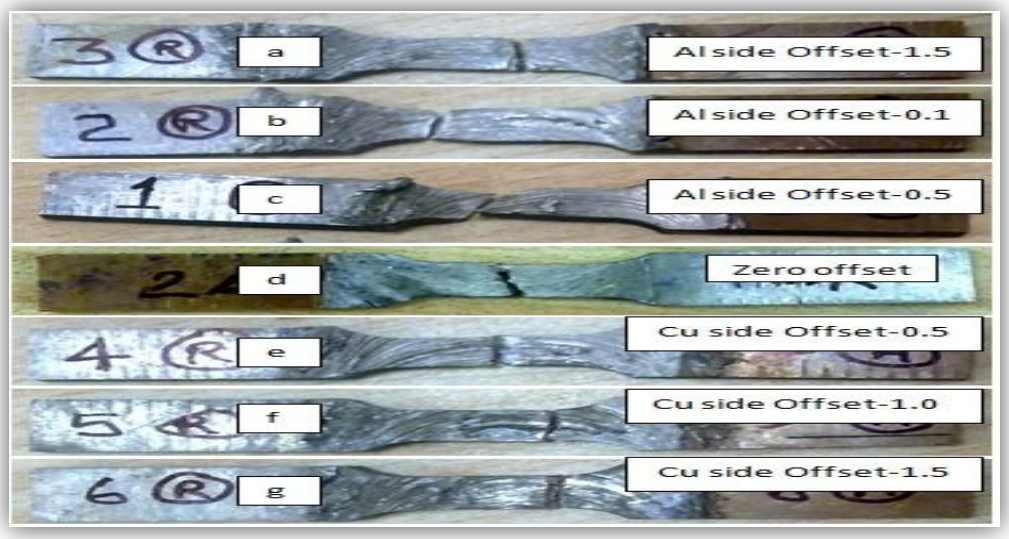


Fig.12. Showing fracture location (a) TMAZ of Cu (b) TMAZ of AL (c) TMAZ of (d)Nugget (e) nugget (f) TMAZ of Cu (g) TMAZ of Cu

Journal of The Gujarat Research Society
ISSN: 0374-8588
Volume 21 Issue 15, December 2019

Fracture location position in the tensile specimens was shown in the fig. 12.while fig.13 shows the variation of tensile strength with tool pin offset. In the present study,tool pin offset was varied between 0 to1.5 mm. Ultimate tensile was found to be 105MPa at zero offset. Tensile strength of the joint was 139MPa at a pin offset of 0.5mm towards Al side. Maximum strength was obtained at a pin offset of 1mm towards Al side. The strength of the joint was decreased at a pin offset of 1.5mm.when pin is offsetting towards copper side then maximum tensile strength was obtained at a pin offset of 0.5 mm and minimum tensile strength was obtained at a pin offset of 1.5mm.

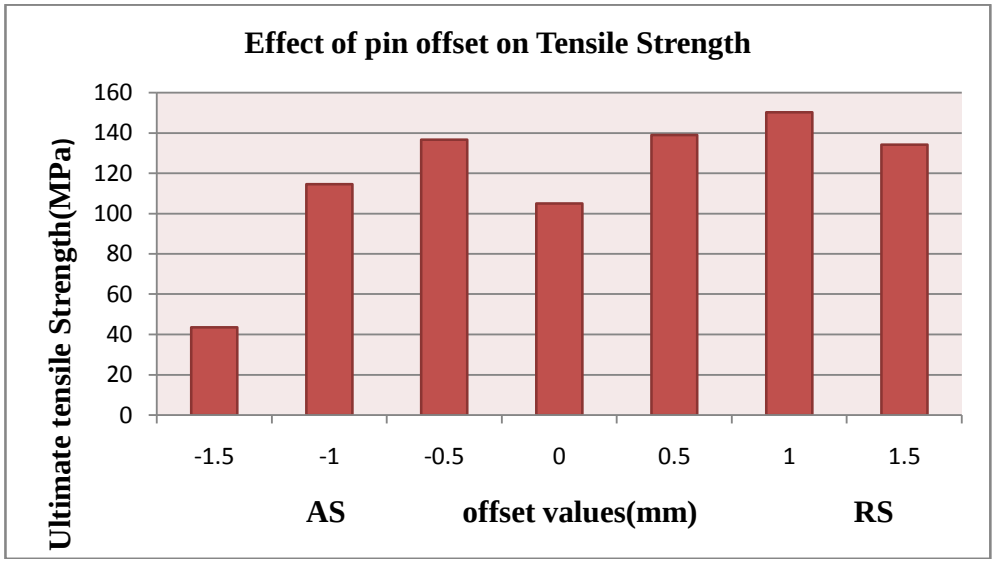
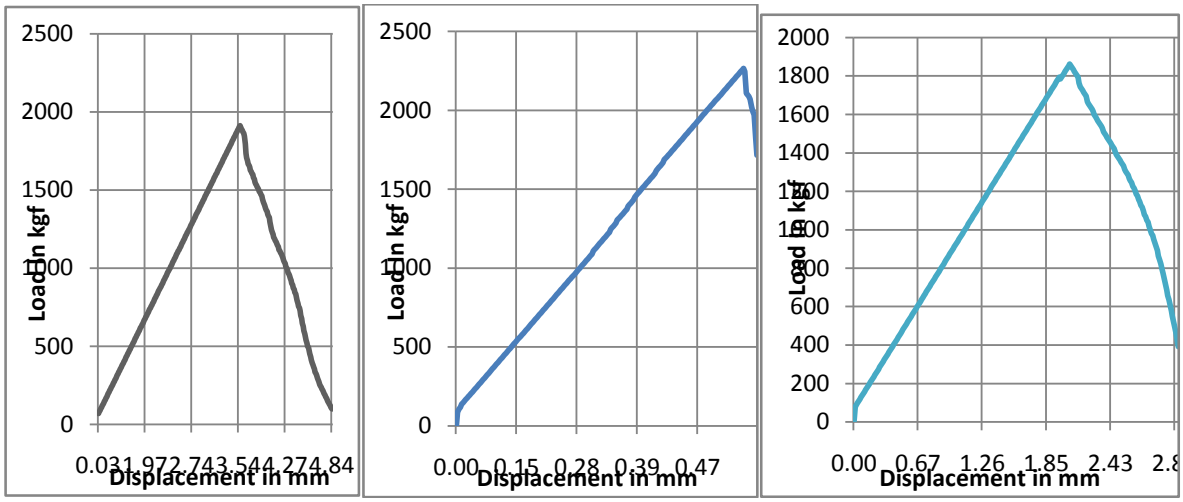


Fig.13. Ultimate tensile strength of welded joint at different offset values**.**

Journal of The Gujarat Research Society
ISSN: 0374-8588
Volume 21 Issue 15, December 2019



(a) (b) (c)

Fig. 14.Stress - strain curves of welded specimens welded with750rpm and 100 mm/min at(a) 0.5 mm tool pin offset, (b) 1mm tool pin offset (c) 1.5 mm tool offset.

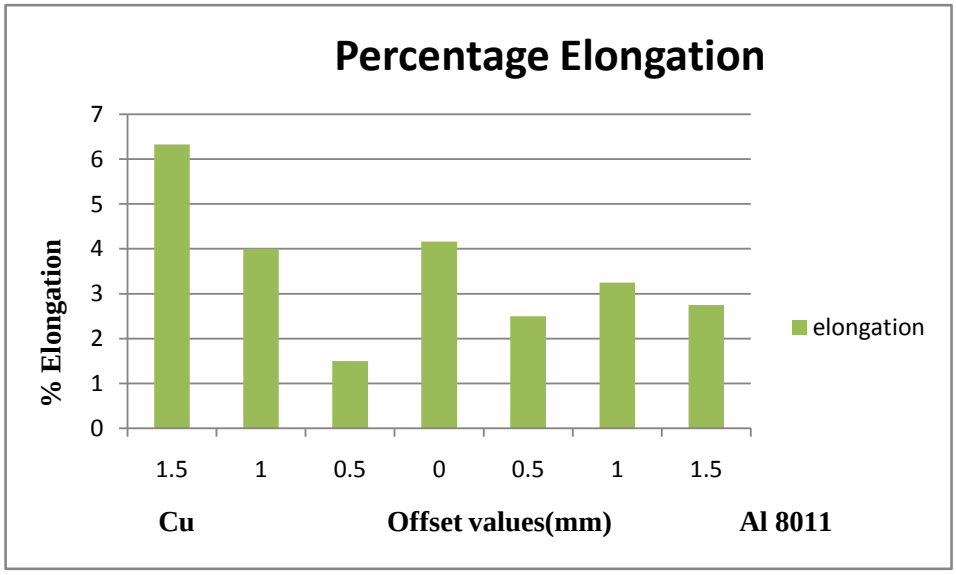


Fig. 15. Percentage Elongation of welded joint at different offset values.

Journal of The Gujarat Research Society
ISSN: 0374-8588
Volume 21 Issue 15, December 2019

3.5 Microstructure analysis

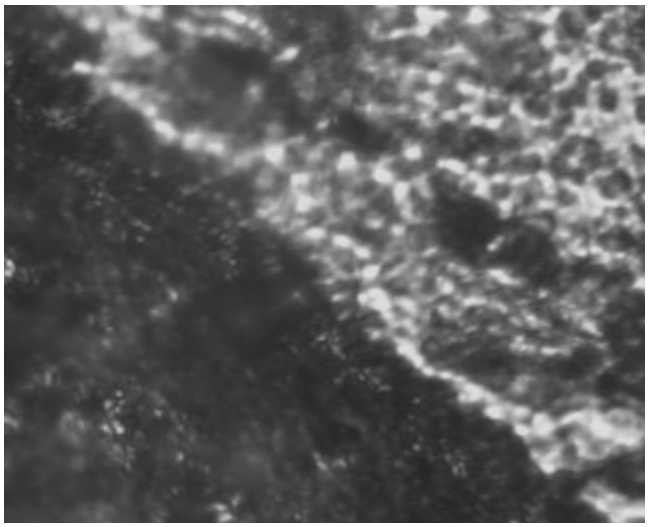


Fig. 16. Microstructure ofAl-Cu Weld interface at750rpm and 100mm/min at 100x magnification

Fig.16shows the microstructure of weld joints at 750rpm and 100mm/min at 100x magnification. The sample was polished using emery sheets of different grades and then disc polished using abrasives, suspended in a water solution, on cloth-covered electrically powered wheeldistilled water. The etchant for Aluminium 8011 alloy is a modified Keller's reagent made of 95ml distilled water, 1ml Hydrofluoric acid, 1.5 ml HCI, and 2.5 ml Nitric acid. For Copper the etchant composition was made of 90 ml distilled water, 5gm Ferric Chloride, and 10 ml Hydrochloric acid. On observing the weld nugget it was seen that the metal close to copper side in the WNZ showed a lamellar structure. However, in the Al side of WNZ the metal showed a mixed structure characteristic. Friction heat of tool induces the different metal structure on both sides. The metal Al experienced a "continuous" dynamic recrystallization (CDRX) and the grain in this region was refined obviously. The metal Cu in weld nugget zone experienced plastic flow by friction heat, but the Cu would not experience the recrystallization process since the temperature developed during welding was much less than the melting point of Cu. As a result, the weld nugget zone with lamellar alternating and mixed structure was formed by stir action and friction heat of tool. Therefore, the Cu and Al were closely combined in the Al side, and showed

Journal of The Gujarat Research Society
ISSN: 0374-8588
Volume 21 Issue 15, December 2019

a mixedstructure.Alternating lamellar structure in weld nugget closeto copper side was shown in fig.17.

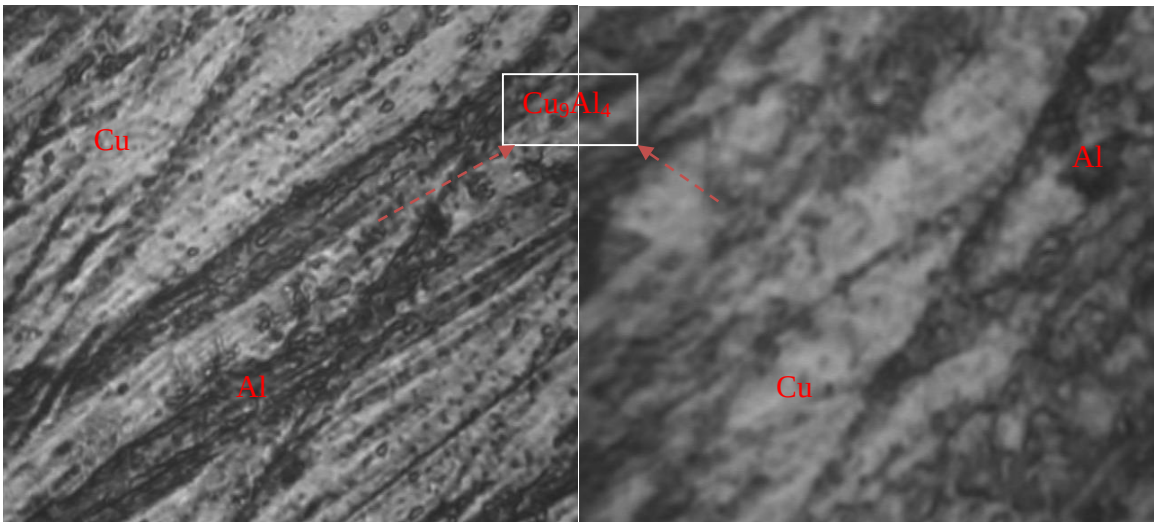


Fig.17. Enlarge view of alternating lamellar structure in weld nugget closeto copper sideat 100x magnification

Fig.18 shows the enlarged views of alternative Cu/Cu_9Al_4 lamellae or vortices that appear near the weld nugget. The bright region are unmixed Cu lamellae with a hardness range of 90-195 $HV_{0.2}$,.The lamellar region appear to be saturated solid solution of Al in Cu and Cu_9Al_4.The concentration of copper at the dark Cu- rich regions has a hardness range of 115-180 $HV0_{.2}$ are considered to contain a certain percentage of the **Cu_9Al_4** intermetallic compound. Similar results were also investigated by Ouyang [17] in his study of dissimilar FSW of aluminum to copper. The formation reason for Cu_9Al_4 is generally attributed tomechanical mixing and interaction in the solid state.

4. Conclusions

AA8011H18 and commercially pure copper were friction stir welded at various welding parameters, andthe effect of the fixed location, pin offset and rotation rate on the microstructure and mechanical properties was investigated. The following conclusions are reached:

1) Sound defect-free joint could be obtained only when the hard Cu plate was fixed at the advancing side. A large volume defectwas observed when the soft Al plate was fixed at the advancing side. This is attributed that the hard Cu bulk material was hardto transport to the advancing side during FSW.

2)Sound defect-free joints were obtained under the pin offsets of 1mm to the Al matrix, and a good metallurgical bonding between the Cu bulk/pieces and Al matrix was achieved. However, defects formed easily at smaller pin offsets due to the hard mixing between the large Cu pieces and Al matrix.

3)Poor tensile properties were obtained at the very large pin offsets and/or low rotation rates due to the insufficient reaction between the Cu bulk pieces and Al matrix. Sufficient reaction were achieved in the FSW Al–Cu joints produced at higher rotation rates and proper pin offset of 1mm towards Al side, resulting in the good tensile properties.

4) The mechanically mixed region in a dissimilar 8011aluminium alloy/copper weld consists mainly of Cu_9Al_4. A mixed layer of Cu_9Al_4 and deformed copper solid solution that showed an intercalated microstruture or vortex flow pattern is formed in copper adjacent to the bottom of the weld by mechanical investigation of Al into copper.

References

[1] Z. Sun,Bonding strength in dissimilar FSWjoints. Int. J. Mater. Prod. Technol. 10 (1995) 16–26

[2] J.H Ouyang, “Material flow and microstructure in the friction stir butt welds of the same and dissimilar aluminum alloys”, pp.110–122, 2006

[3] L.E. Murr, R.D. Flores, O.V. Flores, J.C. McClure, G. Liu, D. Brown, Mater. Res, Innovat. 1, pp.211–223, 1998

[4] W.B. Lee, S.B. Jung, Mater. Res. Innovat. 8, pp.93–96, 2004

[5] P. Vilac¸a, L. Quintino, J.F. dos Santos, R. Zettler, and S. Sheikhi, "Quality Assessment of Friction Stir Welding Joints via an Analytical Thermal Model, iSTIR", Mater. Sci. Eng. A, 445–446, p 501–508, 2007.

[6] D.H. Choi, C.Y. Lee, B.W. Ahn, Y.M. Yeon, S.H.C. Park, Y.S. Sato, H. Kokawa,S.B.Jung, Sci. Technol. Weld. Joining 15, pp.299–304, 2010

[7] T. Saeid, A. Abdollah-zadeh, B. Sazgari, J. Alloys Compd. 490, pp.652–655, 2010

[8] T. DebRoy, H.K.D.H. Bhadeshia, Sci. Technol. Weld. Joining 15, pp.266–270, 2010

[9]T. Watanabe, "Joining of Aluminium alloy to steel by friction stir welding'', pp. 342–349, 2006

[10] P. Xue, B.L. Xiao, D.R. Ni, Z.Y. Ma, Mater. Sci. Eng. A 527 (2010) 5723–5727.

[11] T. Watanabe, H. Takayama, A. Yanagisawa, J. Mater. Process. Technol. 178 (2006)42–349.

[12] T. DebRoy, H.K.D.H. Bhadeshia, Sci. Technol. Weld. Joining 15 (2010) 266–270

[13] J.H. Ouyang, E. Yarrapareddy, R. Kovacevic, J. Mater. Process. Technol. 172 (2006) 110-122

[14] R. Nandan, T. DebRoy, H.K.D.H. Bhadeshia, Prog. Mater. Sci. 53 (2008) 980–1023 A.P. Reynolds, Scripta Mater. 58 (2008) 338–342.

[15] Pratik Agarwal S, Prashanna Nageswaran, N. Arivazhagan & K. DevendranathRamkumar, Development of Friction Stir Welded Butt Jointsof AA 6063 Al Alloy and Pure Copper, International Conference onAdvanced Research inMechanical Engineering (ICARME-2012), 13th May,2012, TRIVENDUM, ISBN: 978-93-81693-59-9

[16] Arbegast. W. J, A flow-partitioned deformation zone model for defect formationduring friction stir welding [J]. Scripta Materialia.

[17] J.H Ouyang, "Material flow and microstructure in the friction stir butt welds of thesame dissimilar aluminum alloys", pp.110–122, 2006

[18] Satyaveer Singh, N.Yuvaraj, Kaleem Uz Zaman Khan, Arshad Noor Siddiquee, Zahid A. Khan, Naveen kumar,"Multi-response Optimization of Friction Stir Welding Process Parameters Using Standard Deviation (SD) Based Preference Indexed Value (PIV) Method",IJAPIE-2018-10-429, Vol 3 (4), 44-48.

[19] Satyaveer Singh, Dr. Zahid A. Khan, Dr. Arshad Noor Siddiquee,"Study on the effect of FSW process parameters on Joint quality of dissimilar materials." Accepted**.** "International Journal of Research in Engineering & Advanced Technology". Volume 3, Issue 2, April-May, 2015 2320 – 8791

[20] Satyaveer singh & Mohd.Atif Wahid,"Effect of tool offset on Friction stir welding of Al alloy to Pure copper"Accepted.Vol.02, Issue No.-08,IJIRSE,August 2016, pp.-330-334

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 12 | Dec 2019 www.irjet.net p-ISSN: 2395-0072

PERFORMANCE ANALYSIS OF C.I. ENGINE USING BIODIESEL FUEL BY MODIFYING INJECTION TIMING AND INJECTION PRESSURE

Ravi Shankar Kumar[1] Dharamveer Singh[2], Desh Bandhu Singh[3]

[1]M.Tech Scholar, Department of Mechanical Engineering, R. D Engg. College Ghaziabad. U.P, India
[2]Research Scholar, Department of Mechanical Engineering, Delhi Technological University, India
[3]Professor, Department of Mechanical Engineering, Galgotia's College of Engg. and Technology, U.P., India

---

Abstract - *This research presents the optimization of injection pressure and injection timing of the compression ignition engine with the use of biodiesel. The fossil fuel or conventional fuel blended with the biodiesel has been done by the fuel modification process. The Jatropha biodiesel blended with diesel for C.I. Engine. B100 (100% biodiesel) biodiesel are used to determine the engine performance on different load. The 2-stroke single cylinder C.I. Engine has been tested with B100 blended fuel. The engine runs on no-load condition for half an hour for the test to obtain the real operational condition of an engine then experiment done. Some blends having high exhaust gases emission such as SOx, NOx but also having high B.P. reading, on the other side some blends having fewer exhaust gases emission of SOx, NOx but also having less B.P. reading. I have made a conclusion that B20 blends are good in all prospect whether it is exhaust gases emission or engine power. I have used Jatropha seed because it will be easily cultivated with less amount of water or in a dry place. The Jatropha plant can evoke in infertile land so that we can cultivate the Jatropha in those types of areas. In the present situation, Jatropha oil are costly as compare to diesel but when the cultivation takes place on the commercial level. The cost will decrease and also do not forget that we have limited conventional fuel for the future generation.*

***Key Words*: Blend[1], CI Engine[2], Emissions[3], B.P.[4], Performance[5]**

1. INTRODUCTION

The energy menace is a situation in which the country experience a break-down of energy supplies brought by quickly increasing energy cost that intimidate the economic and security of a country. The intimidates to economic security is represented by the possibility of retardation in economic growth, increasing inflation, increasing unemployment, and deficit in trillions of money in an investment. The intimidates to country security is represented by the incapability of United State Government to exercise various diplomatics options in particular regarding to nations with substantial crude oil reserves. For example, a few years ago disruption of Venezuelan oil supplies may boundary the US diplomatics options towards Iraq. We have seen the two energy menace of 1973 and 1979, then we feel some similar tenors between the two. Were appertinent with low crude oil stocks, Were appertinent in retarding US petroleum product, Started with political stampede in some of the crude oil producing nations, Were appertinent with high dependency on crude oil imports, Were appertinent with import concentration from a small number of suppliers, Were appertinent with low level of crude oil industry spending, Caused an economic downturn for countries, In Middle East countries limited the United Nation diplomatics, Dismemberment US and punish the west in response to support for Israel in the Yom Kippur war against Egypt led the price of crude to rise

As per the International Energy Agency project report minimum 50% enhancement in demand by 2030. The enhancement in consumption was 3% in 2011, but yearly enhancement of only 1.6% would lead to 51%. India and China are both rapidly enhancing their consumption of old. The supply is not enough to meet the demand and rapidly enhancement in global prices of crude oil is making an impact in the economy of many countries. Therefore it is essential to adjust to the retarding oil supply and turn to an alternative source of energy.

1.1 Dependency on Imported Petroleum Fuel

India is one of the largest crude oil and petroleum product consuming and importing countries. India's rank in oil consumption and import is 3rd and rapidly enhancement in oil consumption the rank go up as per US ministry of Petroleum Planning and Analysis Cell (PPAC) report. India imports nearly 70-72% of its petroleum demands. India had imported 213.83 million tonnes of crude oil 2016-17 for USD 70.236 billion. At present yearly consumption of diesel oil in India is 95 million tonnes constituting nearly 40% of the total petroleum product consumption. In 1990-91, the European Community introduced a 90% tax reduction for the use of biofuels, together with biodiesel. Presently mainly 34 countries worldwide produce biodiesel and India rank's 26th of these. India's crude oil imports go up by 23.4% in July this year on enhancement in demand for oil products, an experienced Petroleum Minister officially said. Crude oil imports in July were 8.532 million tonnes as opposed to 6.943 million tonnes in the same month previous last year.

1.2 World Oil Production

It is manifestly crucial to know the limitation of the world oil reserves. However, it is parallel crucial to know the rate at which oil can be produced. Each and every oil producing

country of the world, scrimp for one, produced oil at almost the maximum rate in 2003, Figure 1 shows it. That mudslinging is Saudi Arabia. Venezuela and Iraq both had production obstructions in 2003 but world oil production remains nearly constant because of Saudi Arabia made up the difference.

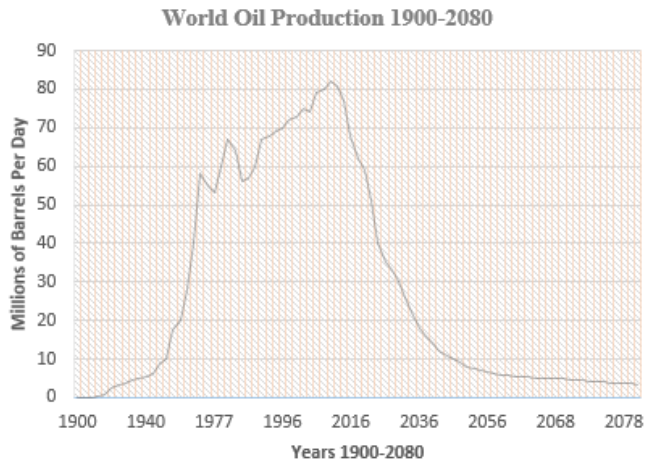


Fig -1: World Oil Production in Millions of Barrels Per Day

Saudi Arabia was able to significantly enhancing production on a short terms foundation in 2003, but perhaps not in 2006, according to Matthew Roy Simmons, the founder and chairman emeritus of "Simmons & Company International" in an oil field. Oil statistics data ordinarily comprise the ratio of R/P (Reserves to Production). It would appear that this is a scale of how long the oil will end. This is an illusory correlation because this assumes that production will remain constant until the last drop is extracted. Here are details of what virtually go on when an oil region is discovered. At first, production proceeds exponentially because of each new oil well adds its output. Lastly, production of the oil region reaches at maximum as every new oil well produces less oil and old oil well run dry. Then there was an exponentially decline as more old oil wells run dry. The productivity of a particular oil well depends on the rate at which oil flows horizontally to the wellhead, the pressure within the oil-bearing creation and viscosity of the oil. There are an optimum rate of production of any well. If oil is extracted too fast then some of the oil is left behind and the quantity ultimately produced is reduced.

1.3 Scope of Study

The further study required to understand the process of making biodiesel from two or more vegetable oils mixture,

[1] Energy and exergy analysis of biodiesels mixture may be conducted and a software programming for chemical composition, energy, and exergy analysis of fuel is to be developed.

[2] Oxidation and storage of the biodiesel fuel properties may be studies.

[3] Development and research have to be taken up for the development of non-edible oil-based lubricants and their commercialization.

[4] Whole life cycle and statistical analysis of individual potential biodiesel feedstock may be conducted.

[5] We required more intensive investigations on exploring different valuable by-products from glycerol and proper utilization of oil cake.

2. LITERATURE REVIEW

High percentage blends of biodiesel can have a problem due to cold temperature. B100 biodiesel produced from soyabean oil will become cloud at temperatures little above freezing and can prevent fuel filters if the temperature drops below 28 ^{0}F. In such conditions, biodiesel blends with diesel fuel are preferred. Since biodiesel is a strong solvent, it will lose trash in tanks and pipes, clogging filters initially. Remedy this problem is after first use change filters. Sometimes gaskets and rubber hoses on older vehicles with B100 don't hold up well. Pre-1991 vehicles should be monitored for hose reduction. If these happen, the seals and hoses should be replaced with Viton- installed parts.

By the use of biodiesel engine warranties are not affected, notwithstanding dealers are continuously confused on this point. The parts and assembly of the engines only cover manufacturers' warranty statements and problems caused by the fuels never cover, regardless of whether the fuel is biodiesel or petroleum-based diesel. The questions arise about compensation for damages caused by a distinct fuel should be addressed to the supplier of fuel.

2.1 What is Biodiesel?

The biodiesel is made from a number of feedstocks together with tallow, lard, vegetables oil and waste edible oils. "Biodiesel is defined as mono-alkyl esters of long chain fatty acids derived from vegetable oils or animal fats. To make biodiesel, feedstocks can be trans-esterification to made biodiesel by using alcohol. Which has been mixed with a catalyst like sodium hydroxide or potassium hydroxide. Most commonly used alcohol for a trans-esterification process is methanol. Methanol is less expensive and easily reacts so it used more than other alcohols.

$$\text{Triglyceride} + \text{Methanol} \xrightarrow[\text{Cat. NaOH}]{} \text{Glycerol} + \text{Esters}$$

(Trans-esterification Reaction)

Biodiesel derived from soybean is generally used in the US. While in Europe generally used feedstock for biodiesel production is rapeseed. Biodiesel is aromatics, biodegradable, nontoxic and essentially free from sulphur, in nature. It is mostly produced from the product that can be grown domestically and internationally, so it is also presumed as a renewable resource.

2.2 Physical & chemical properties of Biodiesel?

Since biodiesel and ethanol are made from biomass materials and its transportation is similar to biofuels. It's commonly used as blended fuel with petroleum fuels namely with gasoline and diesel fuel, but it can also be used alone. Biodiesel and ethanol are generally clean burning fuels, it produced very less air pollution emission.

1. Cetane number: It is defined as the comparative measure of a time interval between the start of injection and autoignition of the fuel. If higher the cetane number then minimum the delay interval and higher the combustibility. The fuel having low cetane number outcomes is difficult to start, also have exhaust smoke and noise. By using a test engine, the cetane number is generally measured and provide knowledge about the diesel fuel ignition quality. Commonly, the cetane number's range between 47 to 50 for diesel fuel, although 40 to 45 for vegetable oils.
2. Melt or pour point: It indicates the temperature at the oil which is in the solid state begin to melt or pour.
3. Cloud point: The temperature at which fuel oil being start solidify is called as the cloud point of that fuel.
4. Carbon residue: This shows the tendency of the fuel to form carbon deposits in an engine. A crucial indicator of the quality of biodiesel is the carbon residue, which coincides to the content of glycerides, liberated fatty acids, polymers, soaps, and remaining catalyst.
5. Density: The weight per unit volume is known as density. The oil having denser contain high energy. By obtained sample values, its seem that biodiesel is normally denser than diesel fuel, varies between 874 kg/m^3 to 881 kg/m^3 compared with diesel which has 848 kg/m^3 density. So, the final product density depends generally on the feedstock used.
6. Kinematic viscosity: Viscosity describes the flow properties and the tendency of a fluid to go under deformation when stress appear. Viscosity makes an impression on fuel atomization and injector lubrication. Fuel having less viscosity, cannot provide enough lubrication for the precision fit of fuel injection pumps, so it's resulting in enhancement in wear or tear and leakage.
7. Calorific value: Heating Value or heat of combustion is the amount of heating energy liberated by the combustion of fuel's unit value.
8. Flash point: The minimum temperature at which the fuel will get fire on an application of a firing source is known as the flash point temperature of diesel fuel. Fresh and clean biodiesel have 140^{0}C flash point, whatever well higher the flash point of petroleum-based diesel fuel ±80^{0}C.
9. Iodine value: It is a listing of the number of double bonds in biodiesel oil, so a parameter that liberates the degree of unsaturation of biodiesel. It is worth of the mass of iodine, dignified in grams, sucked up by 100 grams of specified oil.
10. Acid value: The total acid number is an indication of the presence of free fatty acids formed due to oil degradation and combustion.
11. Stability- oxidative, storage and thermal: Due to the existence of the chemical fabrication of methyl ester and fatty acids in biodiesel, its ages period more hurry as compared to petroleum diesel fuel. Generally, there are fourteen types of fatty acid methyl ester present in the biodiesel. The particular ratio of presence of these esters in the fuel oil makes an impression on the final properties of biodiesel. Lower oxidation stagnation can cause fuel thickening, the formation of adhesive and sediments which in turn can cause separate clogging and injector fouling. Biodiesel or biodiesel blends can't be kept in a storage tank or vehicle tank greater than six months. Depending on the storage temperature and some other circumstance suggest the use of convenient antioxidants.
12. Ash percentage: For the heating value, the ash content is crucial, as ash content enhanced with the retard in heating value. Biodiesel exhausts less ash content as compared to other coals, and sulphur content is too less as compared to other fossil fuels.
13. Sulphur percentage: The percentage by weight, of sulphur in the sulphur containing fuel is limited by law to a very low percentage for diesel fuel used in on-road application vehicles. We use animal fat and vegetable oil-based biodiesel having less than 15 PPM sulphur.

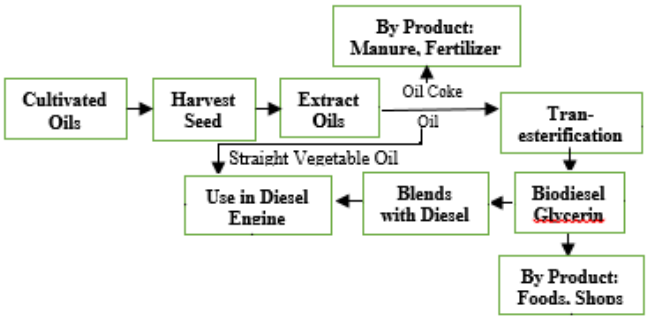


Fig- 2: Schematic Diagram of Biodiesel Production.

Biodiesel is produced by trans-esterification reaction process from animal fats, vegetable oils, and alcohol. Those chemical reaction change animal fats and vegetable oils into a mixture of esters of the fatty acids, which make up the oil from fats. Then the mixture fatty acid methyl esters are purified to obtained biodiesel. A catalyst is also mixture in reaction to enhance the chemical reaction of biodiesel formation. On the basis of the catalyst used in the enhancement of reaction, trans-esterification can be classified into acidic (enzymatic), basic, and former being is most commonly used. A usual trans-esterification reaction is shown in equation RC00RO represent an ester, R0O0H

alcohol, RO0H glycerol (another alcohol), RC00ROO an ester mixture and a catalyst RC00RO Þ R0OOH, cat RO0H Þ RC00ROO. When methanol alcohol used in the trans-esterification process, the produced product of the reaction is a mixture of methyl esters, similarly if ethanol was more important alcohol used in biodiesel production methanol. It most widely used, in spite of its venomousness. It is a liquid material of petrochemical origin. Ethanol fewer used, because it needs more sophisticated production technology and reaction process are lower. It can be easily produced from biomass.

Most of the time trans-esterification reactions is used for biodiesel production at all production scale. For industrial production scale acid less used, occasionally applied the first stage with very acidic raw materials. The enzymatic used very rarely. Lipases are used in the enzymes, the product of the reaction process could be an ethyl esters mixture. In above, both cases glycerin will be the coproduct of the reaction. However, trans-esterification is an extremely crucial step in biodiesel production due to it produces a mixture of esters. Some more steps are needed to obtain a product that follows international standards. In resultant, when the trans-esterification's chemical reaction is completed and the two phases are separated (a mixture of glycerin and esters). To reduce the concentration of agreeable levels the mix of methyl esters must be purified. These comprise the residue of catalyst, methanol, and water; the latter is commonly mixed in standby proportion with the raw materials in order to reach supreme conversion efficiency in the trans-esterification reaction. In the mentioned sections the methods of purification process will be given in details. By treatment of raw materials the content of free fatty acids, moisture and not saponifiable matters are main parameters to obtain high conversion efficiency in the trans-esterification reaction process. The use of available basic catalysts in triglycerides with more content of free fatty acids is not suitable because part of the latter reacts with the catalyst to form soaps. In outcome, part of the catalyst is spent, and no far available for trans-esterification. In short, the efficiency of the reaction reduces with the enhancement of the acidity of the oil. Fundamentals esterification reaction of ethanol with dimethyl carbonate produce glycerol carbonate and methanol. Stages of the biodiesel production process by treatment of raw materials alcohol catalyst mixing chemical reaction dissociation product refinement of the reaction products. Introduction to biodiesel production from vegetable oils, poke, coconut, poultry, animal fats from cattle, an acid trans-esterification is needed a starting stage to retard the level of FFA as per the noted value. Besides having less humidity and acidic material; it is essential that the oil present in a low level of not saponifiable material. If the latter were to exist in considerable amount and soluble in biodiesel. It would lower the level of esters in the product, building it arduous to observe with the less ester content expected by the standards. The American Oil Chemists' Society (AOCS) standards list the expected properties of oils. Anyway, the properties expected by the oils are ultimately determined by the biodiesel industry in every nation. For instance, in Argentina the oils for biodiesel production ordinarily have: Density at 15^0C: 0.70-0.80 g/ml, Moisture content: 450-550 PPM, Peroxide index: 10 mEq/kg, Acidity level: 0.1 mg KOH/g, Non-saponifiable substances: >1%, Kinematic viscosity at 40 ^{0}C: 2-4.3 cSt, Cetane index minimum: 45-55.

Alcohol catalyst blends the alcohol applied for biodiesel production must be mixed with the catalyst before the addition of oil. The mixture is well stirred until the catalyst is perfectly dissolved in the alcohol. It should be noted that for the reasons stated in the last paragraph, alcohol should be water-free (waterless). Sodium and potassium hydroxide is one of the most generally used catalysts. For production on an industrial scale, Sodium or potassium methoxide or methylates are commercially available. Of course, caution is required, and all applicable safety rules Interpretation must be followed, when with methanol, hydrogen and methoxides are independently working on the production scale. Alcohol-to-oil volume ratio, R is another major transformation of the transesterification process. In the stoichiometric ratio, 1 liter of oil is needed to react with 3 liters of alcohol, 3 liters of fatty acids, methyl ester (FAME) and 1 Liter to obtain glycerin. However, since the reaction is reversible, substitute balance on the right side of the additional alcohol equation as a reaction. Which will increase the quantity of products (since it can be estimated from the principle of Le Chatelier). Although a high alcohol-to-oil ratio does not change in the properties of FAME, it is to separate biodiesel from glycerin will make it arduous because it will later enhanced the solubility of the former. Typically, a 100% alcohol intake is used in practice, i.e. 6 mol of alcohol per mol of oil. It concur to 1: 4 alcohol-to-oil volume ratio (R = 0.25). The relationship between alcohol-to-oil ratios, the dielectric properties of FAME, will be discussed later. Lastly, it should be well-known that the essential quantity of catalyst is determined to keeping in mind the acidity of the oil. Biodiesel production action chemical reaction occurs, when the oil needs to be described in the preceding paragraph is mixed with the alkoxide (alcohol-catalyst mixture). It requires certain circumstance of time, temperature and stirring. Due to the high stagnation of the set emulsion, problems arise in the purification of the biodiesel and disjunction, when the use of ethanol in the reaction. A trans-esterification process can be done at different temperatures. For the same reaction time, the conversion is more at high temperature. Because the boiling point of methanol is almost 67^0C (340 K or 153 ^{0}F), the temperature for trans-esterification in atmospheric pressure often occurs in the limit between 48^0C and 58^0C. For paradigm, for the production of biodiesel through rapeseed oil (whose main SFAs are linoleic, palmitic, linolenic, and oleic) and methanol with potassium hydroxide as a catalyst, 64 isomers of triglycerides, 8 monoglycerides, 32 diglycerides can be theoretically possible to find. Their methyl esters, potassium salts of fatty acids, water, potassium meth oxide etc. Reflections on this topic signalize the subsequent general guidelines:

For greater reaction time, the concentration of triglycerides retard the concentration of esters, increases and the concentration of mono-and diglycerides enhanced to a maximum and then retard. Most chemical reactions happen throughout the first minutes.

At the start of the chemical reaction, mono- and the absence of digly (diglycerides) the enhance furthermore retard in their concentration during the position and reaction confirms that ester production from trigly (tricerides) occurs in three platforms, as shown by the subsequent equations:

$$\text{Tri-gly.} + \text{Methanol} \xrightarrow{\text{KOH}} \text{Di-gly.} + \text{Methyl esters}$$

$$\text{Di-gly.} + \text{Methanol} \xrightarrow{\text{KOH}} \text{Mono-gly.} + \text{Methyl esters}$$

$$\text{Mono-gly.} + \text{Methanol} \xrightarrow{\text{KOH}} \text{Glycerine} + \text{Methyl esters}$$

Introduction to Biodiesel Production Several methods, with different levels of equipment complexity and training requirements, have been devised to analyze samples that are mixtures of fatty acids and esters from triglycerides obtained from transesterification of vegetable oils. Where MOH methanol indicates, ME is methyl ester, TG, DG, and MG respectively are tri, DI and monoglyceride, and G is glycerine. The tenors of biodiesel production have been developed to examine the samples with several levels of stuff complexity and training demands, which are a compound of fatty acids and esters from triglycerides, which are acquired from the transformation of vegetable oils.

2.3 Combustion

Combustion in Diesel Engines

For the combustion process that occurs in the diesel engine, there is a relationship between the pressure inside the combustion chamber and the crank angle. Raising the fuel temperature, makes the fuel to ignite immediately, still it is not exposed to a flame. The least temperature that this happens is termed the autogenous ignition point (self-ignition temperature). The fuel is introduced into the combustion chamber and warmed by the air of high temperature and pressure. Then, the fuel self-ignites and burns. In a diesel engine, the fuel ignitability is developed because as the compression ratio rises, the temperature rises suddenly.

Ignition delay period

The interval between the start of fuel injection and the start of combustion in the flaming chamber is associated as the ignition delay period. The opening of diaphragm opens with the change in the tilt on the diagram, or by heat discharge analysis of the data, or in the experimental conditions luminous detector. The opening of the injection can be determined by a needle-lift indicator when the injector needle is removed from its seat. It is more difficult to accurately determine the inside. This is best recognized by changing the tilt of the heat discharge rate from the cylinder discharge data. The engine is well marked in the ignition, IDI engine to recognize in the ignition point is hard.

Premixed combustion stage

Throughout the ignition delay period, the combustion of fuel mixed with air inside the flammability border is accelerated gradually in the crank angle degree - large heat liberation characteristics in this phase. If the fuel received in the combustion chamber during the ignition delay the amount is very high - the pressure in the large heat liberation rate rises immediately which creates the diesel knockdown. For the mean ignition number of the fuel, with a delay in ignition, the expansion stroke happens late in the ignition - deficient combustion, low power formation, poor fuel conversion efficiency. If the pressure gradients are in the scale of 0.4 - 0.5MPa / o CA, the running of the engine is not smooth and Diesel knock starts. This value should be between 0.2 to 0.3MPa / o CA.

Mixing-controlled combustion stage

Once the fuel and air which is incorporated while the ignition delay are consumed, the speed of burn (heat discharge rate) is regulated by the rate at which the mixture becomes ready for burning. The rate of burning at this stage fuel vapour and air mixture are mainly controlled by the process. Liquid fuel atomization, evaporation, pre-flame chemical reactions further affect the rate of heat liberation. Heat wave the rate of price seldom touches to a second peak (which is lighter in magnitude) and next decreases with the graduation of the state. In usual, TDC for low particle emission and high performance (and efficiency). It is acceptable to have a combustion process.

Late combustion stage

Energy discharge rate is resuming at a cheaper rate in the stroke - there are various causes for this: a tiny part of the fuel can't burn still, a part of the energy is being in the burning of burning and fuel-rich products and can be discharged. Cylinder Charge is non-uniform and through this phase the mixture develops a full combustion and light isolated output gases.

2.4 Emission Characteristics

Smoke Emission

The configuration of smoke happens on extreme air lack. The shortage of air or oxygen is being locally inside the diesel engines. As the proportion of the fuel to the air drops, it grows. The test results show that Smoke discharges rise with load increment for all compression proportions, because the development of smoking is strongly reliant on the weight. For three separate compression proportions. There is a change in the smoke discharge for the B20 mix with the load. The smoke values for compression 18 were the smallest of them. Since the engine trying to decrease the emission of

better emission smoke in the raised compression ratio can be inside the cylinder. The values of the smoke of diesel and biodiesel blends in full load performance. For the objective of conducting biodiesel due to atomic oxygen, the value of smoke is reduced, which helps in normal combustion, decreases the smoke.

Un-burnt hydrocarbon (HC)

Another emission result presented by diesel locomotives is UBHC. It includes fuel that is entirely unpublished or only partially burnt. The amount of the UBHC depends on the running conditions of the engine and the fuel attributes. For the different compression ratios, It seems that B20 honge considers the UBHC variation for methyl esters. UBHC emissions were high because the high compression ratio was low Pressure and better combustion can be secured in the temperature and high compression ratio. U-H-17 appearances the UBHC emission for total diesel and biodiesel blends in full load situation. HC discharge was low for biodiesel blends. HC discharges in case the fuel mixture was retard with enhancing quantity of biodiesel. In its molecular structure, the inbuilt oxygen content may be capable for the complete combustion and thus can diminish the HC level.

CO Emission

In the diesel engine, CO intermediate is produced during the combustion phase. But, the diesel engines which operate on the lower side of the emissions of stoichiometric proportion, are smaller. The variation of CO with load for the B20. Emissions on the load will be diminished and it arises for full load as it can be understood from the graph. For the diesel, the emission of CO at full load and simultaneous biodiesel have three separate. The compaction is mixed in compression ratios. CO emissions were lower compared to diesel value for biodiesel services. With the enhanced amount of biodiesel in the emissions, the emission was reduced. The excess amount of oxygen in the biodiesel is suitable for the combustion inside the cylinder, and consequently diminishes CO emissions.

3. EXPERIMENTAL SETUP

B.P.	:	3 kW
MAXIMUM SPEED	:	1500 rpm
NUMBER OF CYLINDER	:	1
TYPE OF IGNITION	:	COMPRESSION IGNITION
METHOD OF STARTING	:	CRANK START
METHOD OF LOADING	:	ELECTRICAL LOADING

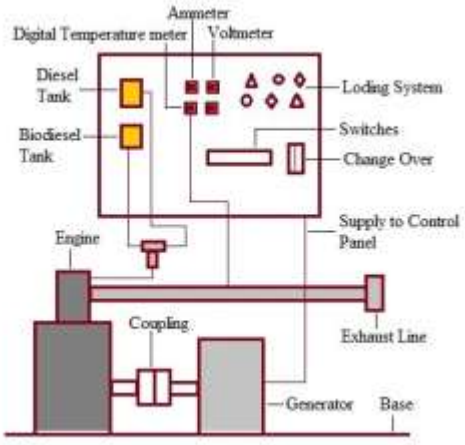


Fig- 3: Schematic diagram of experiment setup

3.1 Loading System

Testing to D.C. Compound is attached to the generator and loaded by an electric bulb. Fuel provided by the central fuel tank through a measuring jar. To prevent the time to stop the fuel consumption of many times to estimate fuel consumption of the engine, fill 10ml fuel the jar by opening the cocked tank several times.

B.P. = (V*I)/ (η_g*1000) kW

η_g = Efficiency of generator.

3.2 Fuel Measurement

The fuel provided from the central fuel tank through a measuring burette with 3 way manifold system. To estimate the fuel consumption of the engine charges the burette by opening the cork marked “tank” in the manifold block, by starting a stopped clock measure the time taken to drain 10 ml of fuel.

m_f = (10*density of fuel*3600)/ (1000*time) kg\hr.

Density of diesel = 848 kg\m^3

Density of biodiesel = 875 kg\m^3

Fig-4: Experimental setup

4. RESULTS AND DISCUSSION

4.1 Observations

Time for 10 ml fuel consumption from jar = t_f , (S)

Exhaust Gas Temperature = E.G.T., (^{0}C)

4.1.1 For High Pressure and Advance Injection Timing

On diesel fuel

Table 1 shows the speed versus load for CI engine with diesel fuel. Fuel injection timing and pressure has a major influence on engine performance, combustion, and emission, and these impacts have mainly executed in case of late injection cases. But, for advance injection timing with high pressure and high heat liberation will attain high values. For homogeneous charging combustion operation engine, compression stroke at advance injection timing and high pressure has been set to ensure better fuel/air mixture quickly. For the end of the ignition, it has electricity 198 V and 1.3 A at 728 rpm while 192 V and 13.8 A at 638 rpm.

Table -1: Summarized various data on diesel fuel

On diesel fuel				
Load (kW)	Speed (RPM)	Vol.(V)	Amp.(I)	t_f(Sec)
0.25	728	198	1.3	108.3435
0.50	723	204	1.6	110.6758
0.75	716	209	2.4	84.7411
1.00	724	213	3.4	75.6188
1.25	712	216	4.6	64.0850
1.50	702	221	5.4	55.4759
1.75	695	215	6.7	47.3648
2.00	677	218	7.7	42.9241
2.25	682	204	8.5	39.0235
2.50	650	198	9.4	35.2992
2.75	664	203	11.4	27.3690
3.00	638	192	13.8	22.0053

On biodiesel fuel

Table 2 shows the speed versus load for CI engine with biodiesel fuel. Fuel injection timing and pressure has a major influence on engine performance, combustion, and emission, and these impacts have mainly executed in case of late injection cases. But, for advance injection timing with high pressure and high heat liberation will attain high values. For homogeneous charging combustion operation engine, compression stroke at advance injection timing and high pressure has been set to ensure better fuel/air mixture quickly. For the end of the ignition, it has electricity 194 V and 1.2 A at 734 rpm while 194 V and 11.4 A at 669 rpm.

Table -2: Summarized various data on biodiesel fuel

On biodiesel fuel				
Load(kW)	Speed(rpm)	Vol.(V)	Imp.(I)	t_f(Sec)
0.25	734	194	1.2	99.5689
0.50	728	203	1.8	84.6215
0.75	722	208	2.5	78.5430
1.00	706	213	3.4	77.0221
1.25	712	218	4.6	64.6261
1.50	705	224	5.8	52.8362
1.75	708	216	6.3	55.0595
2.00	698	206	7.4	51.1388
2.25	691	192	8.7	49.9456
2.50	684	210	9.6	44.6176
2.75	677	206	10.8	34.2358
3.00	669	194	11.4	28.3468

4.1.2 For Normal Pressure and Advance Injection Timing

On diesel fuel

Table 3 shows the speed versus load for CI engine with diesel fuel. Fuel injection timing and pressure has a major influence on engine performance, combustion, and emission, and these impacts have mainly executed in case of late injection cases. But, for advance injection timing with normal pressure and high heat liberation will attain high values. For homogeneous charging combustion operation engine, compression stroke at advance injection timing and normal pressure has been set to ensure better fuel/air mixture quickly. For the end of the ignition, it has electricity 195 V and 1.1 A at 722 rpm while 172 V and 10.2 A at 652 rpm.

Table -3: Summarized various data on diesel fuel

On diesel fuel				
Load (kW)	Speed (RPM)	Vol.(V)	Amp.(I)	t_f(Sec)
0.25	722	195	1.1	168.2820
0.50	712	183	1.7	159.3133
0.75	701	198	2.4	141.5092
1.00	688	205	3.5	110.8045
1.25	692	208	4.6	89.1738
1.50	677	202	5.4	85.7189
1.75	684	212	6.8	62.4419
2.00	671	203	7.4	68.7748
2.25	663	194	8.2	69.5214
2.50	652	186	8.9	70.0011
2.75	643	176	9.4	66.1861
3.00	652	172	10.2	57.3018

On biodiesel fuel

Table 4 shows the speed versus load for CI engine with biodiesel fuel. Fuel injection timing and pressure has a major influence on engine performance, combustion, and emission, and these impacts have mainly executed in case of late injection cases. But, for advance injection timing with normal pressure and high heat liberation will attain high values. For homogeneous charging combustion operation engine, compression stroke at advance injection timing and normal pressure has been set to ensure better fuel/air mixture quickly. For the end of the ignition, it has electricity 196 V and 0.9 A at 718 rpm while 181 V and 10.7 A at 655 rpm.

Table -4: Summarized various data on biodiesel fuel

On biodiesel fuel				
Load(kW)	Speed(rpm)	Vol.(V)	Imp.(I)	t_f(Sec)
0.25	718	196	0.9	96.9150
0.50	712	192	1.4	87.2072
0.75	686	203	2.5	61.3572
1.00	696	205	3.4	52.9973
1.25	689	202	4.8	41.5831
1.50	676	209	5.3	40.0998
1.75	672	197	6.6	35.5884
2.00	648	193	7.4	34.7806
2.25	634	207	8.1	30.7963
2.50	662	194	8.8	31.7150
2.75	642	189	9.5	31.2605
3.00	655	181	10.7	26.3242

4.2 Result

4.2.1 For High Pressure and Advance Injection Timing

On diesel fuel

Table 5 shows the load versus brake power, s.f.c., brake thermal efficiency, and exhaust gas temperature for CI engine with diesel fuel. Fuel injection timing and pressure has a major influence on engine performance, combustion, and emission, and these impacts have mainly executed in case of late injection cases. But, for advance injection timing with high pressure and high heat liberation will attain high values. For homogeneous charging combustion operation engine, compression stroke at advance injection timing and high pressure has been set to ensure better fuel/air mixture quickly. For the end of the ignition, it has B.P. 0.3300 kW, s.f.c. 0.8538 kg/kW-hr, and brake thermal efficiency 9.92% at 0.25 kW load while B.P. 1.8467 kW, s.f.c. 0.3489 kg/kW-hr, and brake thermal efficiency 24.27% at 1.75 kW load.

Table -5: Summarized various data on diesel fuel

On diesel fuel				
Load (kW)	B.P. (kW)	s.f.c. (kg/kW-hr)	B.T.E. (%)	E.G.T. (^{0}C)
0.25	0.3300	0.8538	09.92	233
0.50	0.4184	0.6591	12.85	238
0.75	0.6430	0.5601	15.12	247
1.00	0.9284	0.4348	19.48	255
1.25	1.2738	0.3739	22.65	278
1.50	1.5300	0.3596	23.55	301
1.75	1.8467	0.3489	24.27	319
2.00	2.1520	0.3304	25.63	335
2.25	2.2230	0.3518	24.07	347
2.50	2.3861	0.3624	23.37	361
2.75	2.9669	0.3759	22.53	372
3.00	3.3969	0.4083	20.74	381

On biodiesel fuel

Table 6 shows the load versus brake power, s.f.c., brake thermal efficiency, and exhaust gas temperature for CI engine with biodiesel fuel. Fuel injection timing and pressure has a major influence on engine performance, combustion, and emission, and these impacts have mainly executed in case of late injection cases. But, for advance injection timing with high pressure and high heat liberation will attain high values. For homogeneous charging combustion operation engine, compression stroke at advance injection timing and high pressure has been set to ensure better fuel/air mixture quickly. For the end of the ignition, it has B.P. 0.2984 kW, s.f.c. 1.0599 kg/kW-hr, and brake thermal efficiency 10.63% at 0.25 kW load while B.P. 2.5846 kW, s.f.c. 0.2731 kg/kW-hr, and brake thermal efficiency 41.25% at 2.50 kW load.

Table -6: Summarized various data on biodiesel fuel

On biodiesel fuel				
Load (kW)	B.P. (kW)	s.f.c. (kg/kW-hr)	B.T.E. (%)	E.G.T. (^{0}C)
0.25	0.2984	1.0599	10.63	221
0.50	0.4684	0.7946	14.18	229
0.75	0.6666	0.6015	18.73	239
1.00	0.9284	0.4404	25.58	247
1.25	1.2856	0.3791	29.72	269
1.50	1.6656	0.3579	31.48	288
1.75	1.7446	0.3279	34.36	311
2.00	1.9543	0.3151	35.75	322
2.25	2.1415	0.2945	38.26	334
2.50	2.5846	0.2731	41.25	348
2.75	2.8523	0.3225	34.93	362
3.00	2.8353	0.3919	28.75	373

4.2.2 For Normal Pressure and Advance Injection Timing

On diesel fuel

Table 7 shows the load versus brake power, s.f.c., brake thermal efficiency, and exhaust gas temperature for CI engine with diesel fuel. Fuel injection timing and pressure has a major influence on engine performance, combustion, and emission, and these impacts have mainly executed in case of late injection cases. But, for advance injection timing with normal pressure and high heat liberation will attain high values. For homogeneous charging combustion operation engine, compression stroke at advance injection timing and normal pressure has been set to ensure better fuel/air mixture quickly. For the end of the ignition, it has B.P. 0.2750 kW, s.f.c. 0.6596 kg/kW-hr, and brake thermal efficiency 12.84% at 0.25 kW load while B.P. 2.1223 kW, s.f.c. 0.2054 kg/kW-hr, and brake thermal efficiency 41.22% at 2.50 kW load.

Table -7: Summarized various data on diesel fuel

On diesel fuel				
Load (kW)	B.P. (kW)	s.f.c. (kg/kW-hr)	B.T.E. (%)	E.G.T. (^{0}C)
0.25	0.2750	0.6596	12.84	229
0.50	0.3988	0.4804	17.63	234
0.75	0.6092	0.3541	23.92	244
1.00	0.9198	0.2995	28.28	256
1.25	1.2266	0.2790	30.35	268
1.50	1.3984	0.2546	33.26	281
1.75	1.8482	0.2645	32.02	294
2.00	1.9258	0.2304	36.75	309
2.25	2.0394	0.2153	39.34	327
2.50	2.1223	0.2054	41.22	341
2.75	2.1210	0.2174	38.95	345
3.00	2.2492	0.2368	35.76	354

On biodiesel fuel

Table 8 shows the load versus brake power, s.f.c., brake thermal efficiency, and exhaust gas temperature for CI engine with biodiesel fuel. Fuel injection timing and pressure has a major influence on engine performance, combustion, and emission, and these impacts have mainly executed in case of late injection cases. But, for advance injection timing with normal pressure and high heat liberation will attain high values. For homogeneous charging combustion operation engine, compression stroke at advance injection timing and normal pressure has been set to ensure better fuel/air mixture quickly. For the end of the ignition, it has B.P. 0.2261 kW, s.f.c. 1.4371 kg/kW-hr, and brake thermal efficiency 7.84% at 0.25 kW load while B.P. 2.3019 kW, s.f.c. 0.4377 kg/kW-hr, and brake thermal efficiency 25.74% at 2.75 kW load.

Table -8: Summarized various data on biodiesel fuel

On biodiesel fuel				
Load (kW)	B.P. (kW)	s.f.c. (kg/kW-hr)	B.T.E. (%)	E.G.T. (^{0}C)
0.25	0.2261	1.4371	7.84	224
0.50	0.3446	1.0481	10.75	236
0.75	0.6506	0.7890	14.28	246
1.00	0.8935	0.6651	16.94	258
1.25	1.2430	0.6093	18.49	271
1.50	1.4201	0.5531	20.37	284
1.75	1.6669	0.5309	21.22	294
2.00	1.8310	0.4946	22.78	300
2.25	2.1496	0.4758	23.68	313
2.50	2.1887	0.4537	24.83	328
2.75	2.3019	0.4377	25.74	344
3.00	2.4829	0.4819	23.38	361

4.3 GRAPHS FOR THERMAL EFFICIENCY

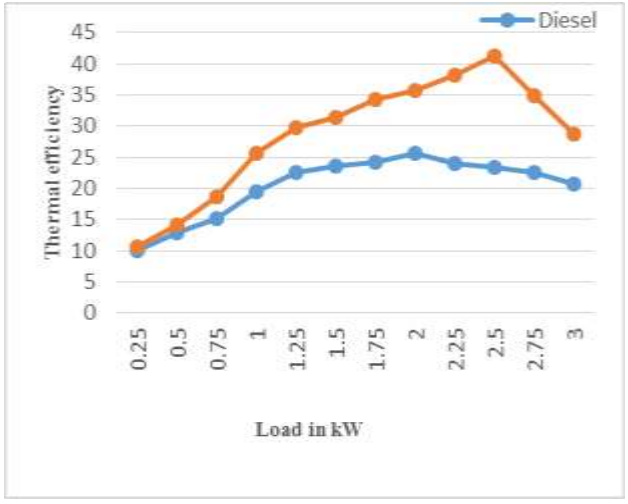


Chart -1: High Pressure Advanced Injection Timing

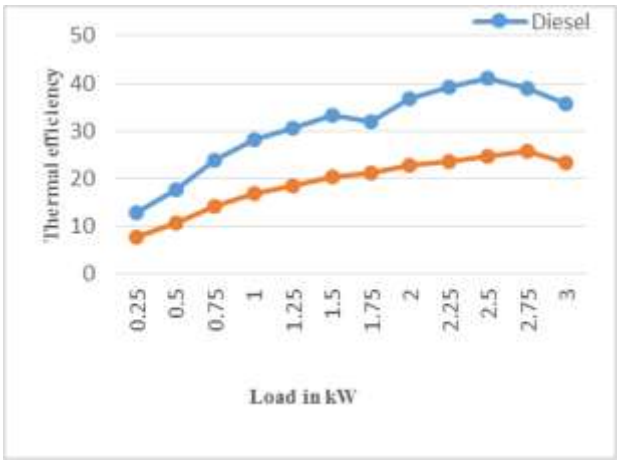


Chart -2: Normal Pressure and Advanced Injection Timing

5. CONCLUSIONS

After examining the record and expertise of those practicing biodiesel, this is desirable that biodiesel is an effective supplement for petroleum diesel for use in heavy vehicles. It provides too many advantages to overlook biodiesel transit agencies. Unlike petroleum diesel, biodiesel is a renewable energy source which is fabricated domestically, which can diminish the reliance on foreign holes and provide more energy assurance. These circumstances singly create biodiesel deserving attention. Comparative experiments were performed to estimate the performance of biodiesel corresponded to diesel on single cylinders compression ignition engine, water cooled. Engine load, engine speed, fuel consumption, and exhaust gas temperature measurement are taken. The result of the experiment can be compiled as follows.

[1] Outcomes confirm that when the fuel is charged with biodiesel, the engine's performance are alike with petroleum diesel.

[2] Fuel consumption with biodiesel is higher than fuel, when fuel is provided with petroleum diesel, because the engine should be provided with a bit more biodiesel to produce the similar amount of work, as before the low calorific value described was done by.

[3] Biodiesel performance in high pressure is compared to diesel.

[4] Combustion properties of the engine confirmed that the highest peak mixture has been shown for diesel after B20.

[5] Better thermal efficiency with high pressure but with advanced timing there is a little more reformation occur.

[6] Performance report of the engine shows that mixes B20 and B50 as a result of high BTE (brake thermal efficiency) corresponded to diesel. The maximum BTE was recorded for the B50, which is 17% more than the diesel.

[7] The brake power for both the fuel examined is comparable. There are little changes with changed parameters.

[8] At advance injection timing and normal injection pressure the thermal efficiency of engine enhanced.

[9] There was slight increase in the broken power of the engine under advance pressure.

[10] When the engine fueled the biodiesel, it would take time to get a steady state.

[11] Exhaust gas temperature (EGT) enhanced when the engine fuels the biodiesel.

[12] In the blends, B20 and B50 produced low CO emissions, compared to Diesel, in B50 a decrease of 34.21% compared to the maximum recorded diesel.

Current test results confirm that biodiesel can be successfully used in existing diesel engines with injection time and injection pressure changes. Biodiesel fuel can be recognized as a competent, engine friendly and environmentally friendly alternative to diesel oil. From the point of view of the social economy. The use of biodiesel a partial diesel option can increase the earnings of farmers. This will also help overcome the uncertainty of the availability of fuel.

REFERENCES

[1] Monteiro MR, Alessandra Regina Pepe Ambrozin, Luciano Morais Lião, Antonio Gilberto Ferreira., (2008) "Critical review on analytical methods for biodiesel characterization." Talanta Vol. 77 pp. 592–605.

[2] Amigun B, Sigamoney R., Blottnitz HV, (2008) 'Commercialisation of biofuel industry in Africa: A review.' Renewable and Sustainable Energy Reviews, Vol.12 pp.689–710.

[3] Agarwal, Avinash Kumar., (2007) "Biofuels (alcohols and biodiesel) applications as fuels for internal combustion engines." Development in Energy and Combustion Science. pp. 230-270.

[4] Agarwal, A K., (2005) "Experimental investigations of the effect of biodiesel utilization on lubricating oil tribology in diesel engines." Prof. IMechE, Vol. 219, Part D: Automobile Engg.

[5] Agarwal Deepak and Agarwal AK., (2007) "Performance and emissions characteristics of Jatropha oil (preheated and blends) in a direct injection compression ignition engine.", Applied Thermal Engg., Vol. 27, pp.2315–2325.

[6] Canakci, Mustafa. Gerpen, Jon H. Van., (2001) "Comparison of engine performance and emissions for petroleum diesel fuel, yellow grease biodiesel, and soybean biodiesel." ASAE Annual International Meeting. No. 016050.

[7] Giibitz G.M, Mittelbach M, Trab M., (1999) "Exploitation of the tropical oil seed plant *Jatropha curcas* L.", Vol.67, pp. 74-82.

[8] Canakci, Mustafa., (2007) "Combustion characteristics of a turbocharger DI compression

ignition engine fueled with petroleum diesel fuels and biodiesel." Bioresource Technology.1169-1175.

[9] Pramanik K., (2007) "Properties and use of jatropha curcas oil and diesel fuel blends in compression ignition engine.", Renewable Energy, Vol.28, pp. 238-247.

[10] Canakci, Mustafa., (2007) "Combustion characteristics of a turbocharger DI compression ignition engine fueled with petroleum diesel fuels and biodiesel." Bioresource Technology.1165-1177.

[11] Forson F.K., Oduro E.K. and Hammond-Donkoh., (2004) "Performance of Jatropha oil blends in a diesel engine.", Renewable Energy ,Vol.29, pp.1133–1147.

[12] Cengel, Yunus A., Boles, Michael A., (2008) "Thermodynamics: An Engineering Approach.", McGraw-Hill.

[13] Agarwal D. and Agarwal A.., (2009) "Performance and emissions characteristics of Jatropha oil (pre-heated and mixture) in a ID compression ignition engine.", Applied Thermal Engg., Vol. 27, pp.2317–2328.

Journal of Optics

ACCEPTED MANUSCRIPT

Petal shaped nanoantenna for solar energy harvesting

To cite this article before publication: RITIKA RANGA *et al* 2020 *J. Opt.* in press https://doi.org/10.1088/2040-8986/ab6ae5

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is **© 2020 IOP Publishing Ltd**.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.
As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by-nc-nd/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 14.139.251.100 on 20/01/2020 at 04:21

IOP Publishing Journal Title
Journal **XX** (XXXX) XXXXXX https://doi.org/XXXX/XXXX

Petal shaped nanoantenna for solar energy harvesting

Ritika Ranga, Yogita Kalra*, Kamal Kishor

TIFAC-Centre of Relevance and Excellence in Fiber Optics and Optical Communication, Delhi Technological University, Delhi-110042

E-mail: dryogitakalra@gmail.com

Abstract

In this article, a petal shaped nanoantenna has been designed and studied for radiation efficiency and harvesting efficiency. Various geometrical parameters such as the length, the width, the height and the gap of the nanoantenna have been varied to find their effect on the radiation efficiency and the harvesting efficiency. Additionally, the significance of the material of the nanoantenna has also been studied by choosing four different types of materials and their corresponding radiation efficiency, harvesting efficiency and electric field enhancement have been reported.

Keywords: solar energy, nanoantenna, plasmonics

1. Introduction

Energy from the sun is the most abundant and freely available energy on the earth's surface. From years photovoltaic (PV) solar cells have been used for solar energy harvesting but there are certain limitations in their use. One of them is their low efficiency. Single PV cell has a conversion efficiency of approximately 29% and multi-junction cells have a conversion efficiency of 46% approximately. One cause of low efficiency is that when light strikes on a PV solar cell, photons having energy higher than the bandgap of the semiconductor are absorbed. The rest of the energy is converted into heat rather than being converted into useful electrical energy. Hence, only a small region of the spectrum is utilized. The efficiency is also affected by input light, surface area, weather conditions, the polarisation of light [1-3]. One of the promising alternatives of a solar cell is nanoantenna that converts electromagnetic light into electrical power over a large bandwidth of the solar spectrum. It covers visible as well as the infrared spectrum. In the nanoantenna surface current is formed which produces electric field between the gap of two nanoantenna. In a metallic nanoantenna, surface plasmons are responsible for field enhancement [4-6].

Nanoantennas can be classified into metallic, dielectric and hybrid nanoantennas [7-9]. The metallic nanoantennas, also known as plasmonic nanoantennas, have been used for energy harvesting applications. At visible and near-infrared (NIR) frequency metals exhibit dielectric properties and allow electromagnetic (EM) radiations to penetrate into them. At the metal surface, when EM radiation of suitable frequency is incident, surface plasmons are formed on the surface of the metal. These surface plasmons produce electric field on the metal surface. The electric field is confined in the gap between the two arms of the nanoantenna. The plasmonic nanoantennas have a wide range of applications in the field of solar energy harvesting [10-12], optical sensors [13], surface-enhanced raman spectroscopy [14,15], optical data storage and optical circuits [16,17] and many more. Further, Bailey et al. had proposed the concept of rectenna in which antenna can be coupled to a rectifier/ diode to convert induced ac power into dc power. In some notable experimental works, the performance of the nanoantennas has been experimentally tested by calculating the extinction coefficient and electric field enhancement [18-22].

In this paper, the solar energy harvesting capability has been investigated by computing radiation efficiency and field enhancement in the optical spectrum. The performance of nanoantenna for energy harvesting largely depends upon the

Accepted Manuscript

 © xxxx IOP Publishing Ltd

radiation efficiency and electric field generated inside the feed gap of the nanoantenna. Hence, the effects of geometrical as well as material parameters on these two characteristics have been discussed extensively. The rest of the article has been divided into four sections, where, section 2 presents the fundamental concept of nanoantenna based solar energy harvesting. Section 3 discusses the design of the proposed structure in terms of geometrical as well as material parameters. In section 4 the technique of numerical analyses and the results thus obtained, have been discussed in detail on the basis of the underlying physics. Finally, the conclusion drawn from the complete analysis has been presented in section 5.

2. The fundamental principle of operation

In terms of design, the proposed nanoantenna, made up of gold, is a structural intermediate of dipole and dimer designs and has been referred to as petal-shaped nanoantenna, on account of its petal shaped arms. Figure 1 shows the schematic design of the proposed nanoantenna indicating the significant geometrical parameters, i.e. antenna arm length l, width w, height h and gap between the two arms g. The proposed structure has been analyzed with respect to all the above-mentioned geometrical parameters.

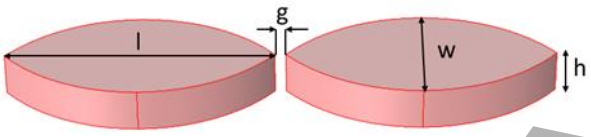


FIG 1. Design of the petal shaped nanoantenna

Further, the performance of the nanoantenna has been analyzed for different materials such as gold (Au), silver (Ag), copper (Cu) and aluminium (Al). The data for the permittivity of materials have been adapted from Johnson and Christy [19]. It should be noted that Drude's model is not applicable because of its inaccuracy in the optical spectrum.

The interest in the optical spectrum is inspired by the fact that most of the solar radiation is concentrated in this region, with its peak value at approximately 500 nm. According to Planck's law of black body radiation, power radiated by a black body as a function of wavelength and temperature is given by

$$P(\lambda,T) = \frac{2hc^2}{\lambda^5}\frac{1}{e^{\frac{hc}{\lambda kT}}-1} \quad (1)$$

bandwidth h is the Planck's constant, k is Boltzmann constant and T is the absolute temperature of the black body, λ is the wavelength and c is the speed of light in free space. Figure 2 shows the solar spectrum i.e. normalized intensity vs wavelength curve computed according to Planck's law by considering sun as a black body with the surface temperature equal to 5777 K. It can be noticed that the spectral peak occurs at 500 nm which corresponds maximum irradiance.

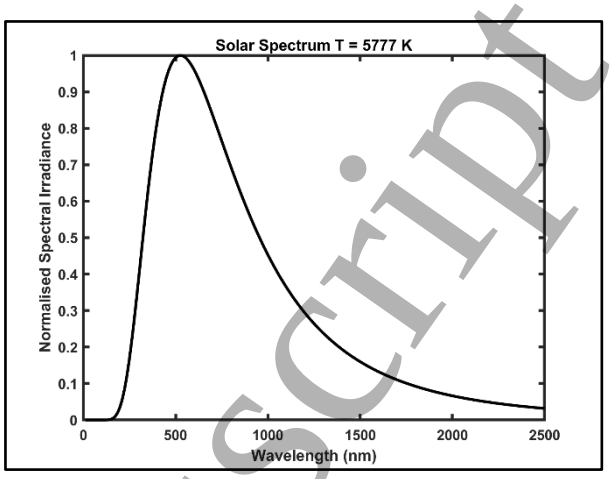


FIG 2: Solar energy spectrum

When the light is made incident on the nanoantenna, some part of the energy is re-radiated and some is absorbed by it. The two most important factors on which the performance of the nanoantenna in solar energy harvesting applications depends are radiation efficiency and electric field enhancement. The radiation efficiency, denoted by η_{rad}, is defined as the ratio of the power radiated by the nanoantenna to the total incident power [1]. Mathematically,

$$\eta_{rad} = \frac{P_{rad}}{P_{rad}+P_{ab}} \quad (2)$$

where P_{rad} is power radiated by the nanoantenna and P_{ab} is the power absorbed by it. P_{rad} can be controlled by varying the size, hence, the effective surface area of the nanoantenna. Greater is the effective surface area, greater is the light gathering capability of the nanoantenna, hence, greater is the radiation efficiency. Radiation efficiency is important because greater is the radiation efficiency greater is the power harvested. However, the total harvested power depends not only on the radiation efficiency but also on the spectral irradiance through the spectrum. Hence, combining Planck's law with radiation efficiency, the harvesting efficiency of the nanoantenna is defined as

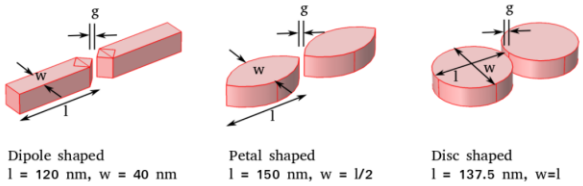
$$\eta_{total} = \frac{\int_0^\infty P(\lambda)\eta_{rad}d\lambda}{\int_0^\infty P(\lambda)d\lambda} \quad (3)$$

The harvesting efficiency is the direct measure of usable output power extracted from radiation falling on the nanoantenna.

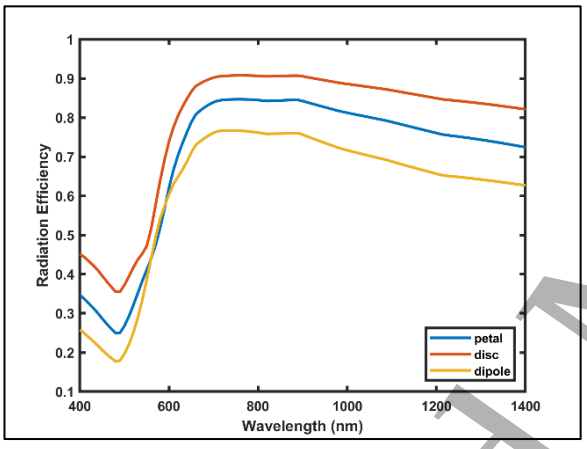
Accepted Manuscript

3. Design and its advantage

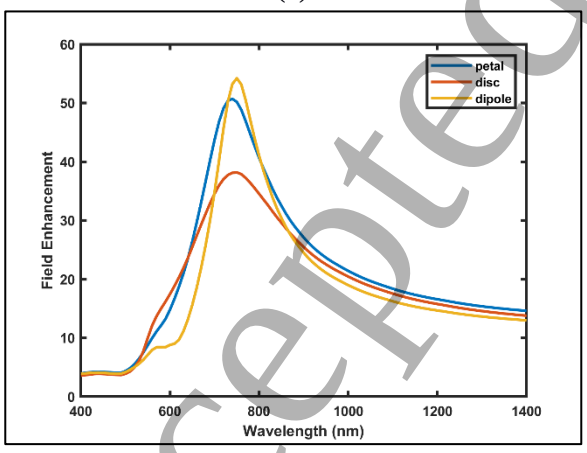
The peculiarity of the petal shape is that it, being an intermediate structure, has partial characteristics of both – dimer and dipole nanoantennas. The following figure illustrates the comparison of the operation of three types of the nanoantenna. In figure 3, a comparative study of electric field enhancement and radiation efficiency has been made between the three designs, which conveys that the dipole antenna has the best field enhancement but lowest radiation efficiency among the three, and vice-versa in case of the disc-shaped nanoantenna. Interestingly, the proposed petal shaped structure exhibits fairly high field enhancement and radiation efficiency.



(a)



(b)



(c)

FIG 3: Comparison of the proposed design with dipole and disc-shaped nanoantennas (a) geometry, (b) Variation of radiation efficiency with wavelength, (c) Variation of Electric field enhancement with wavelength

The high field enhancement is attributed to the pointedness of the structure in the gap region, while high radiation efficiency is attributed to the large surface area offered to the incident radiation. In figure 3, width w is 40 nm, 75 nm and 150 nm in case of the dipole, the petal, and the disc respectively, and length l has been taken to be 120 nm, 150 nm and 137.5 nm respectively. The dipole nanoantenna design has the most pointed edge in the gap region but offers the least surface area to incident light, hence, it provides maximum enhancement but minimum radiation efficiency. On the other extreme end, the disc structure has a round edge but offers maximum surface area, hence, it has the lowest enhancement but maximum efficiency. The petal shaped nanoantenna, being an intermediate design, has pointed edges in the gap region and has greater surface area than a dipole, consequently, its results lie in between the other two and it offers good field enhancement and radiation efficiency. This discourse proves that the proposed structure is a better candidate than the other two.

The proposed nanoantenna has been designed and analyzed by finite element method (FEM) via COMSOL Multiphysics [23]. Plane polarised light is allowed to fall on the nanoantenna with electric field along the direction of length 'l' of the nanoantenna. In all the computations, non-uniform meshing has been used. While an extra-fine mesh, with a minimum element size of 1.5 nm and maximum element size of 35 nm, has been used for modeling of nanoantenna; the surrounding region including the PML has a normal mesh with minimum element size 18 nm and maximum element size of 100 nm. The thickness of the PML has been kept 150 nm. The PML thickness and air gap been the PML and the device has been such chosen that any further increase in either of them does not affect the results anymore. Value for permittivity for gold and other material has been taken from experimental data provided by Johnson and Christy [24].

4. Numerical analysis and results

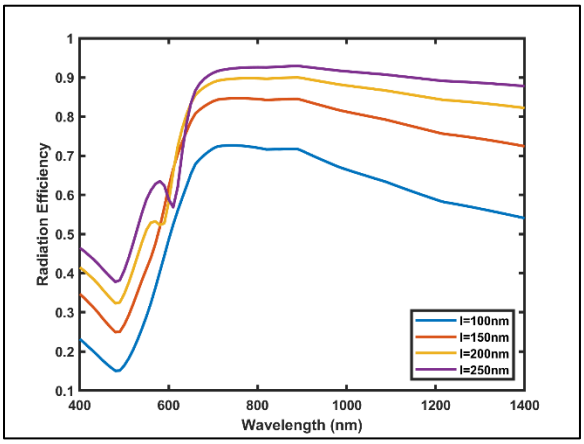
4.1 Effect of length l

Length of the the nanoantenna is the main factor on which the resonant wavelength depends. The structure can be tuned to operate at different wavelengths, simply by changing the length l. In order to study the effect of length, the radiation efficiency and field enhancement have been calculated for different sizes of the nanoantenna. The length has been varied from 100 nm to 250 nm, keeping the width w equal to half of the length (arbitrarily), the height h equal to 40 nm and the gap g equal to 10 nm, in all the cases. The effect of length has been illustrated in figure 4.

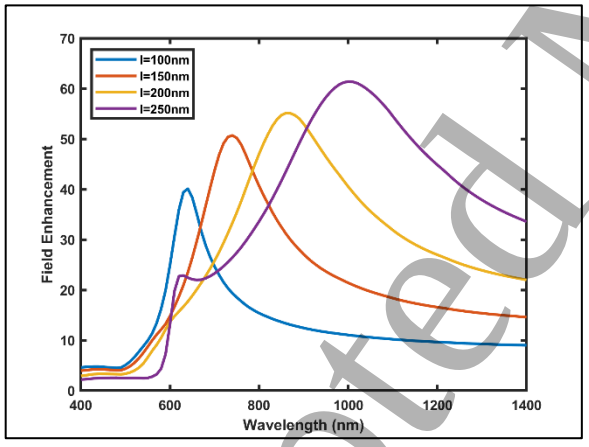
Figure 4(a) represents the variation of radiation efficiency with respect to wavelength for different lengths of the antenna arm. An increase in length results in an increase in effective surface area available for the reception of incident radiation.

Accepted Manuscript

Hence as the length increases the radiation efficiency (or radiation gathering capability) increases, as seen in figure 4(a). On the other side, the increase in length causes a prominent redshift in the field enhancement peak (fig 4(b)), which is a direct consequence of the redshift in the localized surface plasmon resonance upon increasing the dimension parallel to the incident electric field. Maier (2007) gives a detailed account of the plasmon resonance of metallic nanoparticles by presenting polarizability as a function of particle size and its dielectric constant [4]. This implies that the nanoantenna can be tuned to different wavelengths by simply varying the length l. For a conventional dipole antenna, the resonant wavelength is twice its length, but at the nanoscale, this relation does not hold good.



(a)



(b)

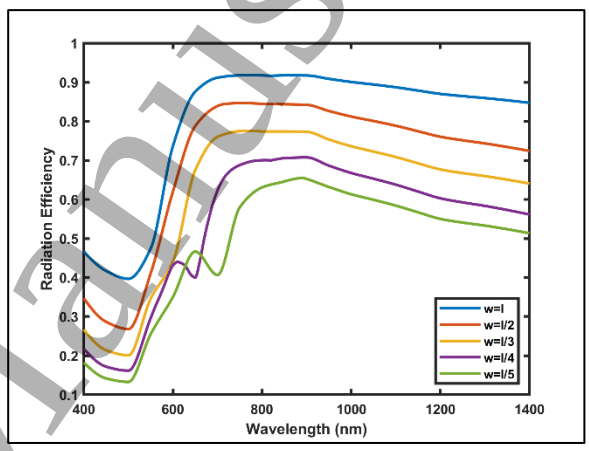
FIG 4. (a) Variation of radiation efficiency and (b) electric field enhancement with wavelength for different length of nanoantenna

TABLE I. Harvesting efficiency for different values of length *l*

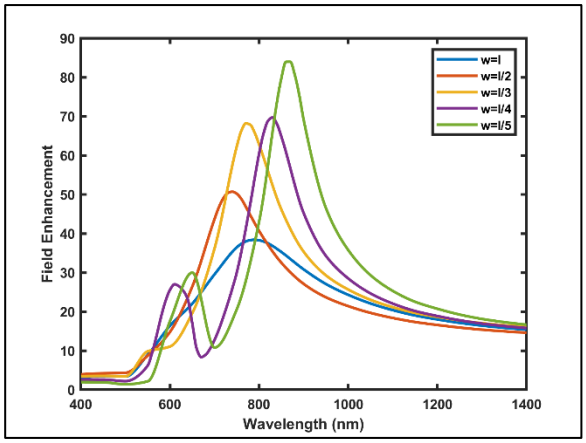
Length (l)	100nm	150nm	200nm	250nm
H.E.	0.5175	0.6484	0.7088	0.7458

4.2 Effect of width w

Besides length, width w is also an important parameter which needs contemplation. Variation in width alters the pointedness of the structure hence directly affects the field enhancement. Effect of width has been studied by setting different values, i.e. w = l, l/2, l/3, l/4, l/5, keeping l constant at 150 nm. When w=l, the structure is simply disc-shaped, whereas, for w=l/5 the antenna arms become so thin that it resembles a dipole nanoantenna. Consequently, when w=l, maximum radiation efficiency (>90%) and minimum field enhancement (~40) is achieved, whereas, for w=l/5, the radiation efficiency is minimum (<70%) and field enhancement is maximum (~85), as shown in figure 5. For all other intermediate values of w, the radiation efficiency and field enhancement lie between the above-mentioned extremities and the structure remains petal shaped.



(a)



(b)

FIG 5. (a) Variation of radiation efficiency and (b) electric field enhancement with wavelength for different width of the nanoantenna.

This analysis exhibits the trend that as the width of the nanoantenna is reduced, keeping the length constant, radiation efficiency decreased due to a reduction in effective surface

Accepted Manuscript

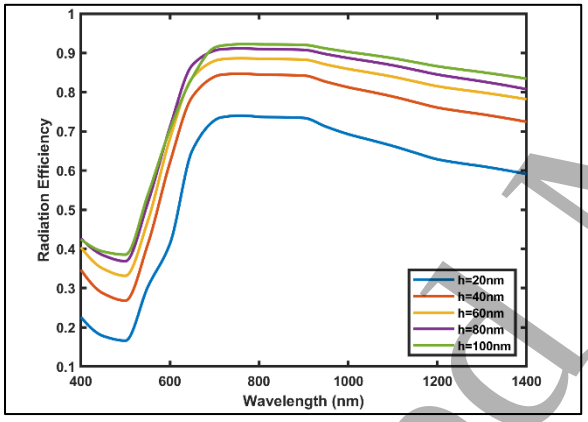
area (as explained in section 4.1), but, field enhancement increases because of the improvement in pointedness (i.e. reduction of tip angle) of the arm. Sharp corners are always the site of near field enhancement due to the lightning rod effect, as explained in rigorous detail by Liao and Wokaun [25]. Hence as the aspect ratio (length/width) becomes more skewed, the corners become sharper and the field at the tip becomes more enhanced [25,26].

TABLE II. Harvesting efficiency for different values of width *w*

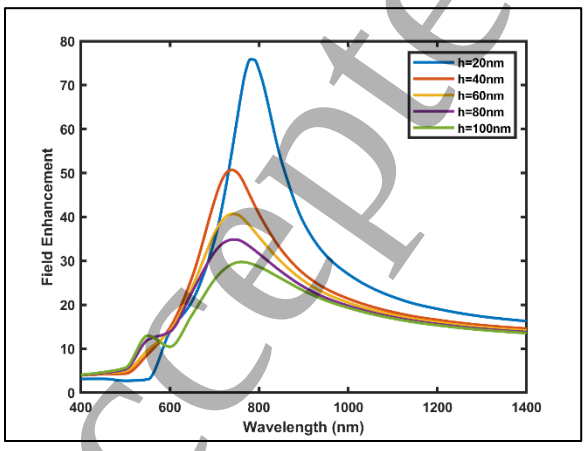
Width (w)	w=l	w=l/2	w=l/3	w=l/4	w=l/5
H.E.	0.7450	0.6497	0.5642	0.4874	0.4285

4.3 Effect of height h

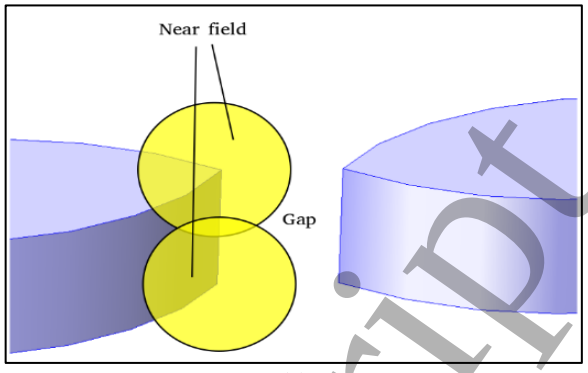
A change in height h of the nanoantenna not only alters its surface area but also controls the field enhancement by altering the extent of overlap between localized surface plasmons at the top and bottom corners of the antenna tip (as shown in figure 6(c)). Hence, as the antenna height increases from 20 nm to 100 nm, though the surface area increases, the overlap of the localized surface plasmons becomes poorer. As a direct consequence, the radiation efficiency increases but field enhancement decreases, substantially.



(a)



(b)



(c)

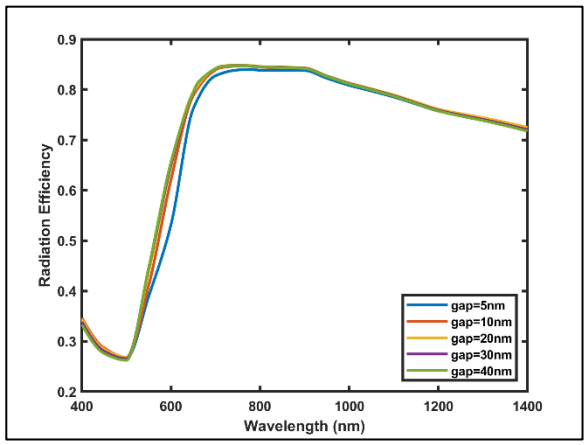
FIG 6. (a) Variation of radiation efficiency and (b) electric field enhancement with respect to wavelength for different value of the height of nanoantenna (c) local field overlap at the corners of single arm of nanoantenna

TABLE III. Harvesting efficiency for different values of height *h*

Height (h)	20 nm	40 nm	60 nm	80 nm	100 nm
H.E.	0.5268	0.6497	0.7012	0.7317	0.7406

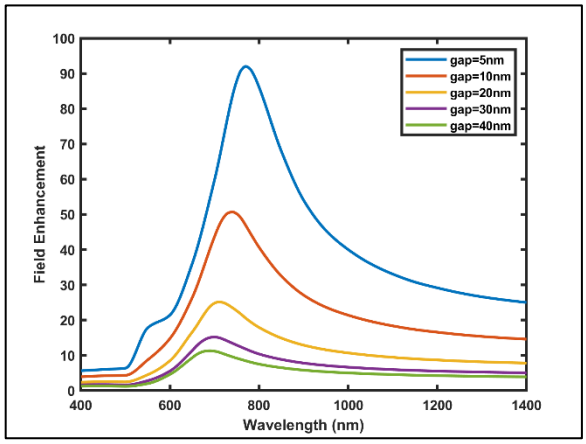
4.4 Effect of gap g

The gap g between the antenna arms is the location of the enhanced electric field, whose magnitude directly depends upon the gap size. The smaller the gap is, the better is the overlap between the localized surface plasmons at tips of the two arms (figure 7(c)), hence, greater is the magnitude of the field enhancement obtained. However, since the surface area is not affected by the change in gap size, the radiation efficiency is practically independent of it. Figure 7 shows the results of numerical analysis in which gap g has been varied from 5 nm to 40 nm. While on one hand, the radiation efficiency remains practically unaffected by it, on the other side, a drastic reduction in field enhancement takes place. This is in absolute agreement with expectations based on the reasoning presented above.

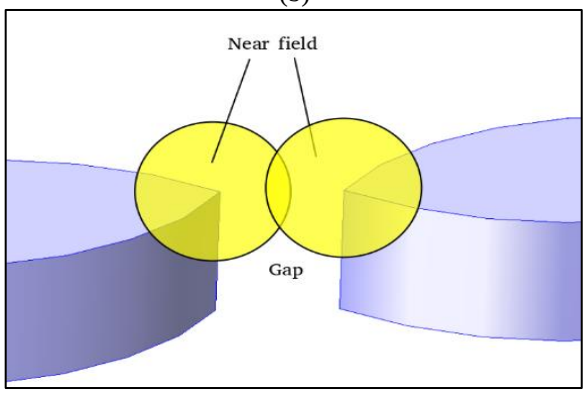


(a)

Accepted Manuscript



(b)



(c)

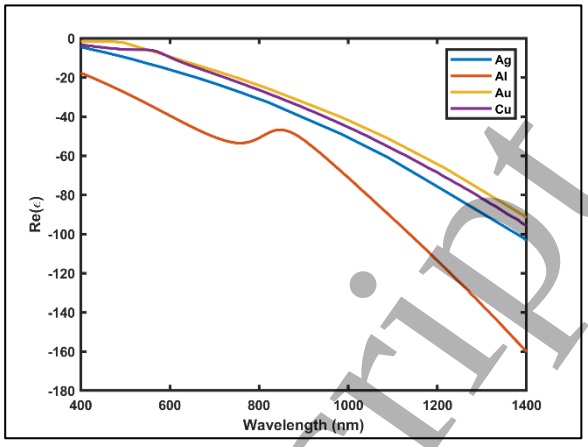
FIG 7 (a)Variation of radiation efficiency and (b) electric field enhancement by varying the tip to tip gap between two nanoantennas (c) local field overlap at the corners of the two arms of nanoantenna

TABLE IV. Harvesting efficiency for different values of gap *g*

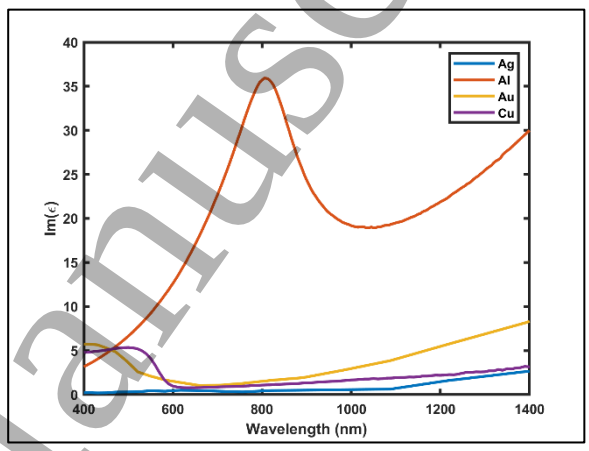
Gap (g)	5 mn	10 nm	20 nm	30 nm	40 nm
H.E	0.6351	0.6497	0.6553	0.6552	0.6545

4.5 Effect of material

The entire analysis presented above has been aimed at demonstrating the effect of geometrical parameters on the operation and performance of the nanoantenna. Besides the geometry, the material of the device is made of, also governs its working. In all the computational results presented in the article so far, gold (Au) has been taken as the material of the device. However, there are other highly conductive materials such as silver (Ag), aluminium (Al) and copper (Cu), whose candidature for device fabrication is worth investigating. Figure 8 below, shows the computational result for the four different materials mentioned above, along with their respective permittivity plots (taken from ref 24).



(a)



(b)

FIG 8. (a) Real and (b) imaginary parts of permittivity of gold, silver, copper, and aluminium

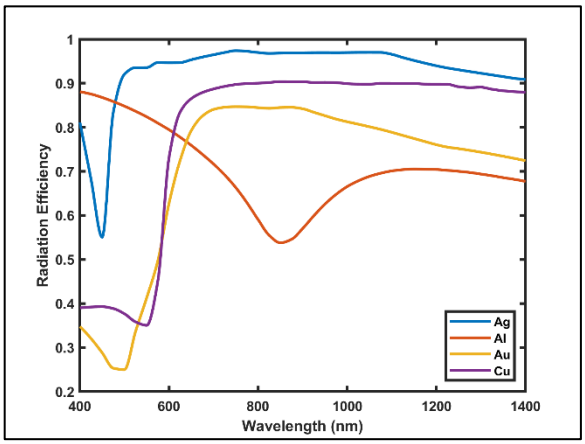
The length, the width, the height and the gap of the nanoantenna, in this case, have been taken to be 150 nm, 75 nm (l/2), 40 nm and 10 nm respectively. Upon observing the graphs for radiation efficiencies it can be inferred that silver is the best performer among all, but its high reactivity makes it unsuitable for long term use.

Gold, which is the most popular material for nanoantenna fabrication due to its chemical inertness, falls short in terms of radiation and harvesting efficiencies. Interestingly, copper and aluminium turn out as moderate performers whose radiation and harvesting efficiencies lie in between the other two.

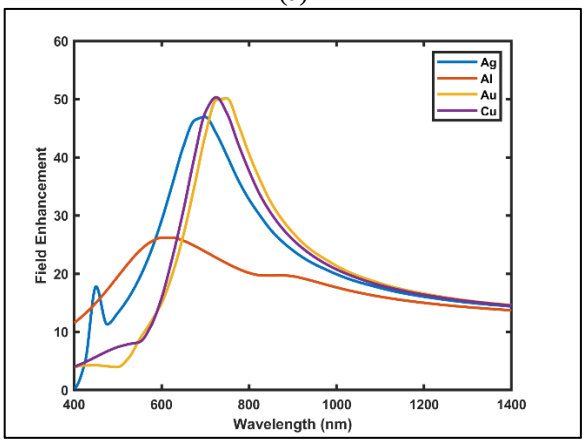
TABLE V. Harvesting efficiency for different materials

Material	Ag	Al	Au	Cu
H.E.	0.9174	0.7312	0.6491	0.7216

Accepted Manuscript



(a)



(b)

FIG 9: (a) Variation of radiation efficiency and (b) Variation of electric field enhancement with respect to wavelength for different materials

5. Conclusion

The proposed design of the petal shaped nanoantenna has been successfully analyzed for solar energy harvesting applications. The two important quantities viz. radiation efficiency and electric field enhancement, have been studied in detail with respect to all the geometrical and material parameters. Furthermore, a comparative study of the proposed design with a well-acknowledged dipole nanoantenna has been performed, in which, the proposed structure turns out to be a better candidate for solar energy harvesting.

Acknowledgements

The authors gratefully acknowledge the initiatives and support provided by the establishment of the TIFAC-Centre of Relevance and Excellence in Fiber Optics and Optical Communication at Delhi Technological University (Formerly Delhi College of Engineering) Delhi, through the "Mission REACH" program of Technology Vision-2020 of the Government of India. One of the authors, Ritika Ranga, expresses gratitude towards her colleague Nishant Shankhwar for his valuable suggestions.

References

[1] Krasnok A. E, Maksymov I. S., Denisyuk A. I., P. A Belov, A. E. Miroshnichenko, Simovski C. R., and Kivshar Y. S. 2013 *Physics-Uspekhi* **56(6)** 539

[2] Gosciniak J. and Rasras M. 2019 *OpticsExpress* **27(6)** 8605-8611

[3] Green A. M. 2015 *Progress in photovoltaics: research and applications* **23(1)** 1-9

[4] Maier S.A. 2017 Plasmonics: fundamentals and applications. Springer Science & Business Media

[5] Zheng P., Kasani S. and Wu N. 2019 *nanoscale horizons* **4** 516-525

[6] Pors A., Albrektsen O., Radko I. P. and Bozhevolnyi S. I. 2013 *scientific reports* **3** 2155

[7] Shankhwar N., Kalra Y., and Sinha R. K. 2017 *Superlattices and Microstructures* **111** 754-759

[8] Reena, Kalra Y., Kumar A. and Sinha R. K., 2016 *Journal of Applied Physics* **119(24)** 243102

[9] Rusak E., Staude I., Decker M., Sautter J., Miroshnichenko A. E., Powell D. A. and Kivshar Y. S. 2014 *Applied Physics Letters* **105(22)** 221109

[10] El-Toukhy Y. M., Hussein M., Hameed M. F. O., Heikal A. M., Abd-Elrazzak M. M., and Obayya S. S. A. 2016 *Optics express* **24(14)** A1107-A1122

[11] El-Toukhy Y. M., Hussein M., Hameed M. F. O., Obayya S. S. A. 2018 *Plasmonics* **13(2)** 503-510.

[12] Moddel, Garret, and S. Grover, eds. "Rectenna solar cells", 4, New York: Springer, (2013).

[13] Mubarak M., Sidek O., Abdel-Rahman M., Mustaffa M., Mustapa A. K. and Mukras S. 2018 *Sensors* **18(11)** 3714

[14] Neubrech F., Huck C., Weber K., Pucci A. and Giessen H. 2017 *Chemical Reviews* **117(7)** 5110-5145

[15] Nien L.W., Lin S. C., Chao B. K., Chen M. J., Li J. H., and Hsueh C. H. 2013 *The Journal of Physical Chemistry* C **117(47)** 25004-25011

[16] Gosciniak J. and Mahmoud 2019 *optics express* **27(6)** 1508-1517

[17] Shiono T., Matsuzaki K., and Furumiya S. 2013 *optics express* **21** 8605

[18] Gadalla N.M. 2013 "Nano Antenna Integrated Diode (Rectenna) For Infrared Energy Harvesting," Diss.

[19] Gadalla M. N., Abdel-Rahman M. and Shamim A. 2014 *Scientific reports* **4** 4270

[20] Ding W., Bachelot R., Kostcheev S., Royer P., and de Lamaestre R. E. 2010 *journal of applied optics* **108**, 124314

[21] Chen S. W., Huang Y. H., Chao B. K., Hsueh C. H. and Li J. H. 2014 *Nanoscale Research Letters* **9** 681

[22] Hussein M., Areed N. F. F, Hameed M. F. O. and Obayya S. S. A. 2014 *Iet Optoelectronics* **8(4)** 167-173

[23] C. Multiphysics® COMSOL AB, Stockholm, Sweden

[24] Johnson P. B. and Christy R.W. 1972 *Physical review B* **6(12)** 4370 (1972)

[25] Liao P. F. and Wokaun A. 1982 *the journal of chemical physics* **76** 751

Accepted Manuscript

[26] Sturman B., Podivilov E. and Gorkunov M. 2013 *Europhysics letters* **101** 57009

Accepted Manuscript

Pragmatic Insights: Live UIs for Agent-based Modelling in Pharo

Jigyasa Grover
Delhi Technological University (formerly Delhi College of Engineering) India
grover.jigyasa1@gmail.com

Nick Papoulias
IRD, Sorbonne Universités, UPMC Univ Paris 06, Unité Mixte Internationale de Modélisation Mathématique et Informatiques des Systèmes Complexes (UMMISCO) , 32 Avenue Henri Varagnat, 93143 Bondy Cedex, France
npapoylias@gmail.com

Serge Stinckwich
IRD, Sorbonne Universités, UPMC Univ Paris 06, Unité Mixte Internationale de Modélisation Mathématique et Informatiques des Systèmes Complexes (UMMISCO) , 32 Avenue Henri Varagnat, 93143 Bondy Cedex, France
Université de Caen Normandie, Caen, France
serge.stinckwich@ird.fr

Pierre Bommel
UR GREEN, CIRAD
Brasilia Brazil
bommel@cirad.fr

Abstract

Object-oriented programming has had a long-standing history with simulation systems in terms of human-computer interaction [1] dating back to Simula and early versions of Smalltalk-72 and Smalltalk-76. These frameworks were based on discrete, event-based simulations as revealed by the classic simulation examples of Smalltalk-80 Blue-book [2]. Nevertheless present time trends have pivoted towards agent-based modelling and abstractions, with systems like NetLogo capturing a sizeable following and heed. Two of the widely known agent-based modelling (ABM) platforms in Smalltalk (MobiDyC and CORMAS) have recently begun experimenting with porting their systems to Pharo. The rationale behind this choice is to allow modern re-imagination of the platforms that can take advantage of developments in: agile visualization, moldable tools, domain-specific languages (DSLs) and so on. This work focuses on the salient user interface component of these platforms, namely the spatial interface, which empowers a user to visualize the evolution of the system through time. We start with a concise analysis of the state-of-the-art for spatial interfaces and their platforms (MobiDyC, NetLogo, GAMA and CORMAS), scrutinizing in terms of programming flexibility, extensibility, portability, scalability, and interaction. Subsequently we highlight our on-going efforts in building an open-source extension for CORMAS' spatial interface in Pharo. In closing, we detail the pragmatic experience gained in the course of our work and present future perspectives.

Keywords Agent-based Modelling; Simulation; Spatial Interface; MobiDyC; NetLogo; GAMA; CORMAS; Pharo; Open Source

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
IWST '17, September 4–8, 2017, Maribor, Slovenia.
Copyright © 2017 ACM ISBN 978-1-4503-5554-4/17/09…$15.00.
http://dx.doi.org/10.1145/10.1145/3139903.3139915

1. Introduction

Object-orientation originated alongside simulation systems in Simula, owing to changes in programming culture and advances in programming technology. With unified concepts facilitating the creation of discrete event systems, programming languages like Simula and early versions of Smalltalk-72 and Smalltalk-76 have provided an across the board foundation for simulation systems. Latter-day modelling systems established on the principal constituent of hierarchy of classes representing agents topped with utilitarian methods have emerged to furnish potent techniques for exploration and evaluation of intricate spatial systems. Common uses of ABM and simulations comprise social or biological system simulations such as crowd behaviour, urban simulation or disease propagation. Encompassing the elements of game theory, complex emergence, computational sociology and evolutionary programming, these ABM systems predominantly include intuitive spatial interfaces to vividly visualize and partially control the evolution of the systems they describe.

With temporal concerns getting complex with each passing point in time, ABM has manifested as a fine computational modelling mechanism enabling the simulation of actions of heterogeneous entities and interactions between them in a collective spatial environment. ABMs are conceptualized as a micro-scale model [3] that simulates the simultaneous operations and reciprocal influences of multiple agents in an attempt to re-create and predict the appearance of the complex phenomena in question. ABMs thus provide explicit views to follow each agent individually and to understand the consequences of multiple interactions in the system as a whole [4].

Many modern ABMs supporting visually robust spatial user interfaces, NetLogo for instance, have been widely adopted both for research and education purposes. Simplistic ABMs like NetLogo make it easier to dip into the simulation sphere with sparse programming skills yet leading to quality outcomes by exploring connections between micro-level behaviors of individuals and macro-level patterns that emerge from their interactions [5]. Dynamically evolving, graphically lucid and adaptive tool-kits with functionalities dispensing live-exploration, time saving simulation rendering along-with enhanced scalability are looked forward to in this era of multiplexity along with the flexibility to mould the source code.

RIGHTSLINK

Hence, from the perspective of Smalltalk primarily, ABMs namely MobiDyC and CORMAS have shifted focus to open-source technologies. The grounds behind this choice is to allow modern reimagination of the platforms that can take advantage of developments in: agile visualization, moldable tools and DSLs.

However, contemporary implementations of ABMs do exhibit impediments in live explorations, programming flexibility extensibility, portability, scalability, heterogeneous data support and automation. In this regard we present a pragmatic report to investigate in detail certain properties of these platforms (MobiDyC, NetLogo, GAMA and CORMAS). Subsequently we focus on our ongoing efforts in developing a live spatial interface for CORMAS (http://cormas.cirad.fr) in Pharo.

Further in this study, Section 2 presents a short present-day state of ABMs, over-viewing properties of the afore-mentioned platforms and related work in this context. In Section 3, the motivation and objectives behind the porting of CORMAS to the open-source Pharo platform is put forth. We then detail the implementation and progress of our spatial interface for CORMAS, considering the case-study of the ECEC Model [6]. Finally, Section 4 sums up presenting a road-map to subsequent work.

2. Problem Statement & Related Work

With ABM toolkits of diverse nature [7] available worldwide, the choice of which one is best suited for a project is left to word of mouth, past experiences, and toolkit publicity apart from their compendious features and aesthetic aspects. One of the most important components of these systems is their spatial user interface especially for end-users, which seeks to facilitate the implementation of modelling systems thus allowing the users to interact with the simulations in real-time. [8]. This involves a quest for dynamic evolvability, adaptive graphical interfaces along with automation and run-time generation for enhanced scalability with efficient patterns materializing in systems completely described by simple rules based on local information.

In this section a concise analysis is presented to study the focal characteristics of a handful of agent-based modelling systems namely MobiDyC, NetLogo, GAMA and CORMAS and also shed some light on limitations in terms of programming flexibility, extensibility and more.

2.1 MobiDyC: Individual Based Modeling for Community Dynamics

MobiDyC is a modelling support multi-platform tool developed in Cincom® VisualWorks® for research in the field of ecology, biology and environment. It has been initiated as part of the working group "dynamics of stands and multi-agent systems" [9]. In MobiDyC, an agent's behaviour is specified either by stringing out pre-defined components or by conventional programming. An important characteristic is the usage of a dedicated scheduler with support for graphical tools, to manage and control simulations in parallel and sequentially [10].

MobiDyC has a very flexible architecture and is suitable for non-computer expert end-users. It helps define agents as autonomous objects that perceive and react to their environment in sync. It focuses more on what each agent does than on what it actually is, thus giving importance to the behavior of individuals. These behaviors are clustered into a low number of classes of activities, e.g. locate, select, translate, compute, end, and workflow control [11].

Despite the many positive aspects, MobiDyC is a very domain-specific platform and its underlying simulation engine is tailor-made for form-based programming [12]. The platform is specifically dedicated to population dynamics with 2D-discrete spatial representation [11]. A further study [13] indicates the following curtailments:

- Discrete space and time may limit the choice of time steps, mesh and communication between agents
- Low numerical precision
- Non-optimized calculations that limit the size to around 10,000 cells and 10,000 agents

The aforementioned points indicate that though MobiDyC has proven to be extremely useful and appropriate for initial model design for education purposes, but the limitations hinder its usage for more realistic simulations [13].

2.2 NetLogo

NetLogo is a popular multi-agent programming language and modeling environment for simulating complex natural and social phenomena. Written in Scala and Java which runs on the Java Virtual Machine, its core is a hybrid interpreter/compiler that partially compiles user code to JVM bytecode [14] thus enhancing portability and also reducing the size of the application.

NetLogo is particularly well suited for modeling complex systems evolving over time [15]. Being a free and open source software, under a GPL license, NetLogo dispenses a flexibility to the end-users to modify it according to the needs and thus contribute to the growth of the tool.

The NetLogo environment enables exploration of emergent social phenomena in a vivid manner by modifying switches, sliders, choosers, inputs, and other interface elements. It is packaged with an extensive library including models of a variety of domains, such as economics, biology, physics, chemistry, psychology, system dynamics, social networks, etc [16]. Beyond exploration and in-depth investigation, NetLogo allows authoring of new models and modification of existing models lucidly.

While NetLogo offers a dedicated programming language that facilitates the implementation of simple models, its scripting language is not object-oriented and does not enable to define subclasses of *Turtle* (one of NetLogo's main abstractions; NetLogo represents its agents as turtles) for example, thus rendering it inextensible in many cases. It is also difficult to deal with data-heavy models, like the ones relying on complex Geographic Information System (GIS) data [17]. NetLogo provides limited support for reading/writing binary files and object-oriented features like data hiding and class inheritance. However, it does allow the user to visualize 3D simulations in limited capacities but not in a consistent manner as it requires language extension and 3D graphic engines for easier builds [18].

Finally, NetLogo's in-built experiment automator, allows separate instances of the same experiment to be run on multiple cores, but it is not yet possible to speed up NetLogo by spreading out the work of a single simulation across multiple cores [19].

2.3 GAMA: GIS and Agent-based Modeling Architecture

GAMA is a Java based open-source modeling and simulation development environment for building spatially explicit ABMs [20]. It is popular owing to the vital characteristics such as high-level and intuitive agent-based GAMA Modeling Language (GAML), data-driven models, declarative user interface supporting deep inspections, multi-layer 2D/3D displays and multiple application domains.

GAMA is extremely handy for 3D modelling for indoor and outdoor spatial environment together with the inclusion of 3D visualizations of agent simulations [18]. The GAMA platform is commonly used in the environmental sector application due to its specialty to integrate GIS with agent-based modeling and multi-scale

RIGHTSLINK

ABM Systems	Pros	Cons
MobiDyC	Multi-platform tool Flexible architecture Elementary GUI for beginners	Enterprise tool Domain specific Space & time computation constraints
NetLogo	Free & Open Source environment Lucid extensibility with third-party apps Partial compilation to JVM bytecode	Single instance, single core Inability to model information rich systems Limited object-oriented support
GAMA	Open source tool Support for 3D modelling & visualization Easy partnership with third-party apps	Complex GUI Support for spatially explicit ABMs Specific to GIS data models
CORMAS	Open-source & free software Based on MVC architecture Detailed sensitivity analysis Comprehensive spatial interfaces	Based on proprietary language Limited extensibility with third-party apps Opaque model structure

Table 1. Comparative study of the different ABM platforms viz. MobiDyC, NetLogo, GAMA and CORMAS.

control. As with NetLogo, GAMA performs modeling and authoring environment with its syntax called GAML where GAMA simulation engine provides the ability to support multi-level models.

The graphical user interface of GAMA is quite complex mainly for beginner users. GAMA provides assorted models libraries for the user to probe topped with the special agent behaviours (called *skills* in GAMA). This marked feature allows users to call their own functions or even enable advanced users to create their own skills into their codes [20]. GAMA similar to NetLogo is GPL licensed thus allowing users to experiment and bring out different colors of the tool partnering with different third-party extensions. Nevertheless as in the case of NetLogo the DSL of GAMA lacks in both abstractions and supporting tools when compared to an OO environment.

2.4 CORMAS: COmmon-pool Resources and Multi-Agent Simulations

CORMAS is an ABM tool based on the Cincom® VisualWorks® platform dedicated to natural and common-pool resources management [21]. CORMAS is intended to facilitate the design, monitoring and analysis of ABM scenarios. As an open-source and free software (MIT license), it is used by an international community of researchers willing to understand the relationships between societies and their environment [22]. In the purpose to deal with natural renewable resources management, CORMAS is mainly oriented towards the representation of interactions between stakeholders and their environment with a specific focus on the interactions between natural and social dynamics [23].

CORMAS complies with the Model–View–Controller (MVC) architecture allowing the modeller to focus solely on the subject, thus freeing him/her from the innumerous coding constraints. The model thus designed is independent of any perspective, different Point of Views (PoVs) could be stipulated to spot the entities [23] but evidently there could be limitations owing to opaque model structure [24]. Probes come across as unique to CORMAS which are specified explicitly by the modeller. Several boosted interfaces are provided within easy reach to model the display and curves of the probes.

CORMAS also supplies a credible feature of running multiple simulations at once along with the three important types of sensitivity analyses: *simple stochastic analysis* that repeats several simulations, *OAT analysis (One factor At a Time)* to study the signature of the parameters (the value of a parameter is gradually or randomly changed for each simulation), and *crossed analysis* for which several parameters' values are changed simultaneously. Similar analysis can also be performed via R (programming language).

Lately CORMAS has taken an innovative direction oriented towards collaborative design of models and interactive simulations with hybrid ABM systems being developed with inputs from stakeholders and end-users. CORMAS has also been enhanced with a feature that enables the customization of spatial interface, plugging in with third-party apps for enhanced visualizations or enabling advanced human-computer mechanisms that fit specific ways to perceive and interact with the entities represented in the model being simulated [23]. It also provides a pronounced facet to act directly on the space and also on the agents during the simulation. This allows users to interact with the running simulation by simply changing the behavior of agents. In CORMAS, a simulation can be distributed on several machines, and several users can remotely manipulate entities and also monitor the same simulation with different viewpoints. Thus, it is possible to collectively explore medium and long-term scenarios to better understand how a desired situation may be reached. In turn, this feeds back into the collective design of the model.

Nevertheless, there is an inherent trade-off between using an agent-based DSL and a general purpose language for modelling and simulation. On one hand using the full syntax and semantics of Smalltalk for example (as CORMAS does) is advantageous both in terms of available object-oriented abstractions and for tool support (browsers, debuggers, profilers etc). On the other hand using such a general-purpose object-oriented language presents a large learning curve for students and experts outside informatics. CORMAS also inevidently fosters opaque model structures [24]. Furthermore while the core of CORMAS itself provides a limited fluent API for modelling, the same is not true for tool or visualization extensions. We believe that it is precisely in this context that the methodologies of agile visualization [25], moldable tools [26] and embedded DSLs [27] can prove useful.

Summing up, we present a synopsis of the comparative study of the different ABM platforms viz. MobiDyc, NetLogo, GAMA and CORMAS in Table 1.

3. Our spatial interface for CORMAS in Pharo

As a framework that proposes predefined classes and a set of visualization tools, the CORMAS environment is intended to facilitate the implementation of ABM systems as well as simulation monitoring and analysis. CORMAS currently uses Cincom® Visu-

RIGHTSLINK

alWorks® (VW) [28], a proprietary cross-platform programming environment based on Smalltalk. Our effort is to make CORMAS fully open-source by porting the modelling software from VW to Pharo. While doing so we are trying to recreate the platform itself, by taking advantage of modern technological advancements in Pharo like the Roassal visualization engine [25], the Spec UI framework [29], the Glamorous Toolkit [30], *PetitParser* [31] and others. In this work, we mainly focus on porting the spatial interface of CORMAS in Pharo.

3.1 Implementing a live spatial interface in Pharo

We start our implementation by supporting the visualization of a CORMAS model instance inside *RTView*, a container of graphical elements representing a layer that may be combined with other views, in Roassal visualization engine without using any additional UI widget library. Subsequently in order to support supplementary interactions, we embed this Roassal view inside a Spec UI view and also use Glamorous Toolkit (GTools) for building a customized inspector.

Roassal Visualization Engine Roassal [25] is a visualization engine predominantly used within the Pharo community to visualize large sets of data. An instance of CORMAS Model is initialized in *loadSimulationWithModel:* method in a *RTGridLayout* (Line 18) which places the elements in a grid. The cells of the grid are defined as a *RTBox* (Line 10) each of a given dimension, here *25* (Line 12). Each cell element is further given a different shade (of green, in this case) which varies according to the attribute value. This is achieved by implementing *RTMultiLinearColor* and using *RTColorPalette* (Line 6).

A look at Figure 3 shows a running simulation with the cells of the CORMAS Model grid simulation changing colors in real-time with shades being dependent on an attribute, for example on *biomass* in the ECEC Model (described in Section 3.2) .

```
1  loadSimulationWithModel: aCORMASModel
2    |n c|
3    aModel := aCORMASModel
4      initialize;
5      new.
6    n := RTMultiLinearColor new colors: (RTColorPalette
       ↪  sequential colors: 9 scheme: 'YlGnBu').
7    c := n command: [ :cells | cells biomass / 10.0 ].
8    aModel initSimulation.
9    aModel runStep.
10   elements := (RTBox new
11     color: c;
12     size: 25) elementsOn: aModel
         ↪  theCMECECVegetationUnits.
13   view addAll: elements.
14
15   "Adding a popup for biomass info on hover"
16   self onHoverPopup: elements.
17
18   RTGridLayout new
19     gapSize: 0;
20     lineItemsCount: aModel spaceModel column;
21     on: view elements.
22   view canvas camera focusOnCenterScaled.
```

Spec UI Framework The Spec UI framework [29] is used in Pharo to define a layout for all the widgets in the user interface. Here, *defaultSpec* is used as a pre-selected method which puts into a columnar layout (Line 3-8) the three components: Menu or the toolbar built using Spec *MenuModel*, Diagram containing the CORMAS Model from Roassal *RTView* and a Spec *TabManagerModel* which gathers multiple tabs and makes them work in sync.

```
1  defaultSpec
2    <spec: #default>
3    ^ SpecLayout composed
4      newColumn: [ :col |
5        col
6          add: #menu height: self toolbarHeight;
7          add: #diagram;
8          add: #manager height: 100 ];
9        yourself.
```

Initially, a *SpecToolbar* is set up for the ECEC model specifically by adding a basic window with a CORMAS Model diagram and a *MenuModel* with load option. With flow, the *Load* menu in the toolbar is detailed by allowing loading of different CORMAS Models including the ECEC Model, the Fire Automata Model, and so on as shown in Figure 5 by populating the menu statically for now. A separate execution panel is created (refer *executionPanelMenuGroupOn:* method) with working *Run* (Lines 2-7), *Stop* (Lines 8-13) and *Reset* (Lines 14-19) buttons. The previously developed user interface is polished by adding icon images for buttons in the spec toolbar (Line 6, 12 and 18).

```
1  executionPanelMenuGroupOn: aMenuGroupModel
2    executionPanelMenuGroup
3      addItem: [ :anItem |
4        anItem
5          name: 'Run';
6          icon: (self iconNamed: #smallDoItIcon);
7          action: [ self animateSimulation ] ].
8    executionPanelMenuGroup
9      addItem: [ :anItem |
10       anItem
11         name: 'Stop';
12         icon: (self iconNamed: #smallWarningIcon);
13         action: [ aModel simManager finishSimulation ]
             ↪  ].
14   executionPanelMenuGroup
15     addItem: [ :anItem |
16       anItem
17         name: 'Reset';
18         icon: (self iconNamed: #refreshIcon);
19         action: [ self resetSimulation ] ].
```

A *Visualize* option is affixed in the *SpecToolbar* to help picture various stages of simulation and the data collected correspondingly, refer Figure 4 for instance. Also, the visualization sub-menu is made dynamic in nature to make it more generic and independent of the model used by selecting methods based on their protocols (Line 8 in *visualizationSubMenu* method). A sub-menu is also developed under the *Visualize* option to provide possibilities to picturize data for various entities and PoVs of the model over the course of time. A method similar to visualization is created for entities to view their information while the simulation is running with dynamic sub-menu where attributes are collected via the *modalClasses* dictionary. As can be referred by observing the space user interface built so far depicting running simulation and probe visualizations in Figure 4, rudimentary simultaneous views of the running system are already supported. We intend to extend this feature with the generic infrastructure of Roassal, called *builder*, to efficiently build and reuse visualizations tailor made for each domain [32].

Sample tab features is added at the bottom of the Spec window using the *TabManagerModel* to display entity classes and their corresponding methods in near future.

```
1  visualizationSubMenu
2    | attributeSelectors |
3    ^ MenuModel new
4      addGroup: [ :group |
5        modelClasses
6          ifNotNil: [ modelClasses
```

RIGHTSLINK

```
7               keysDo: [ :aModelClass |
8                 attributeSelectors := self
                    ↪  collectAtttributeSelectors: (self
                    ↪  class environment at: aModelClass)
                    ↪  withTag: #accessing.
9                 attributeSelectors
10                  ifNotNil: [ :anAttributeSelectors |
11                    group
12                      addItem: [ :item |
13                        item
14                          name: anAttributeSelectors
                             ↪  asString;
15                          icon: (self iconNamed:
                             ↪  #nautilusIcon);
16                          action: [ self visualize: self
                             ↪  aModel on:
                             ↪  #CMECECVegetationUnit with:
                             ↪  #biomass ] ] ] ] ] ]
```

The Glamorous Toolkit *GTInspector* from the GTools [30] moldable development environment, which provides multiple interchangeable presentations for every object is utilized to build customized inspectors initiated on *onClick* and *onHover* events for the entities in the model to highlight only relevant information as depicted in Figure 7 and Figure 6 respectively. Furthermore, the customized *GTInspector* is made independent of the model utilized by directly extending it in *CORMAS Core* package in the super-class *SpatialEntityCell*. This is done as each child-class for the entity cell inherited from the super-class *SpatialEntityCell* implements its own methods for *id*, *state*, *occupants*, etc. and thus shows specific values corresponding to the child-class instance. *gtInspectorCustomization: composite* method is defined in the *CORMAS Core* package in the super-class *SpatialEntityCell* and a composite table (Line 3) is used to display the key-value pairs for attributes like id, state, current time, coordinates, neighbourhood, occupants (Lines 6-12) and so on.

```
1  gtInspectorCustomization: composite
2    <gtInspectorPresentationOrder: -10>
3    ^ composite table
4      title: 'Spatial Entity Cell';
5      display: [ {
6        ('id' -> self id asString).
7        ('self' -> self asString).
8        ('state' -> self state asString).
9        ('current time' -> CormasModel timeStep asString).
10       ('coordinates' -> self coordinates asString).
11       ('neighbourhood' -> self orderedNeighbourhood
          ↪  asString).
12       ('occupants' -> self theOccupants asString).
13       } ];
14     column: 'Key' evaluated: #key;
15     column: 'Value' evaluated: #value;
16     send: #value
```

The UML diagram of the CORMAS Pharo package is depicted in figure 1 showing chief instance variables and methods. It can be observed that the main class *CORMASPharoUI* inherits from *ComposableModel* provided by the Spec UI framework. It uses *RTBox*, *RTdraggableView* and *RTAnimation* from the Roassal visualization engine to view the different *CORMASModel* sample projects. It further uses *MenuModel*, *MenuGroupModel* and *TabManagerModel* from the Spec UI framework to fabricate a robust graphic user interface. To include custom inspectors initiated by *onHover* and *onClick* events, *GtInspector* from GTools is being utilized on the *SpatialEntityCell* class from *CORMAS Core* package to facilitate customization.

Using Roassal as our visualization sub-system, Spec framework for composable UI and GTools for custom inspectors has demonstrated to provide a good base for overcoming the short-comings of

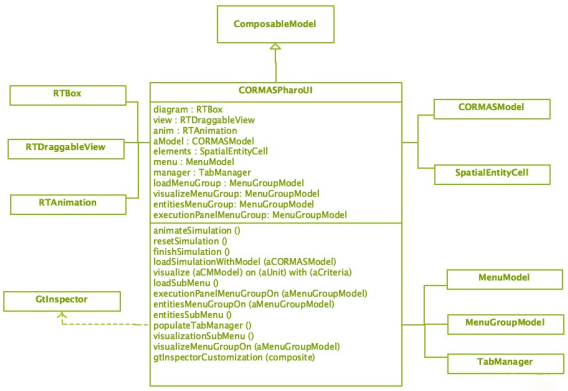


Figure 1. UML diagram of the CORMAS Pharo package

present day enterprise natured ABMs in an Open Source environment. In course of collaborative and agile development of the proposed live spatial user interface extension of CORMAS in Pharo we were able to leverage the above mentioned frameworks to support multiple visualization perspectives, run-time probing of simulations along-with providing appreciable programming flexibility.

3.2 Case Study: The ECEC Model

The ECEC model presented as one of the sample projects in CORMAS, is inspired from a paper by Pepper and Smuts, "Evolution of Cooperation in an Ecological Context" [6]. This model consists of a two dimensional grid, wrapped in both axes to avoid edge effects. It contains two kinds of entities: plants and foragers. The main idea is the study of the survival of two populations of foragers that depends on the biomass availability of the plants and also of the initial spatial configuration of the landscape.

Figure 2 represents the different components of the ECEC Model as included in CORMAS consisting of restrained foragers, unrestrained foragers and plants with chief attributes like *harvestRate*, *biomass*, *energy*, *catabolicRate* and methods like *step()*, *energyConsumed()*, *reproduce()*, etc.

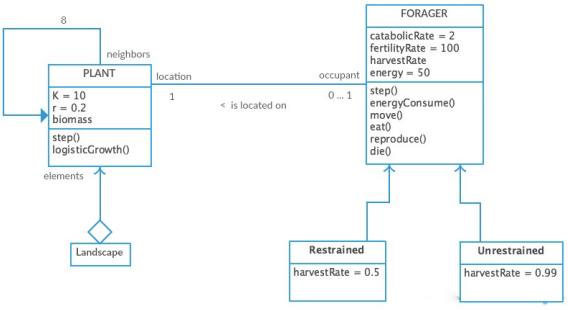


Figure 2. UML diagram of the ECEC Model of CORMAS [33]

The present Pharo extension of CORMAS administered on the ECEC model can be run by executing the script *self new openWithSpec* on *CORMAS Pharo* package and selecting the *CMECEC Model* from the drop-down load menu in the toolbar at the top of the window that opens up as shown in Figure 5.The expected output of a running simulation of ECEC Model can be seen in Figure 3. By hovering on each cell in the grid-layout, we can see the value of an important attribute of that particular cell as in Figure 6. On clicking, a customized inspector with details of that particular cell comes up as shown in Figure 7.

RIGHTSLINK

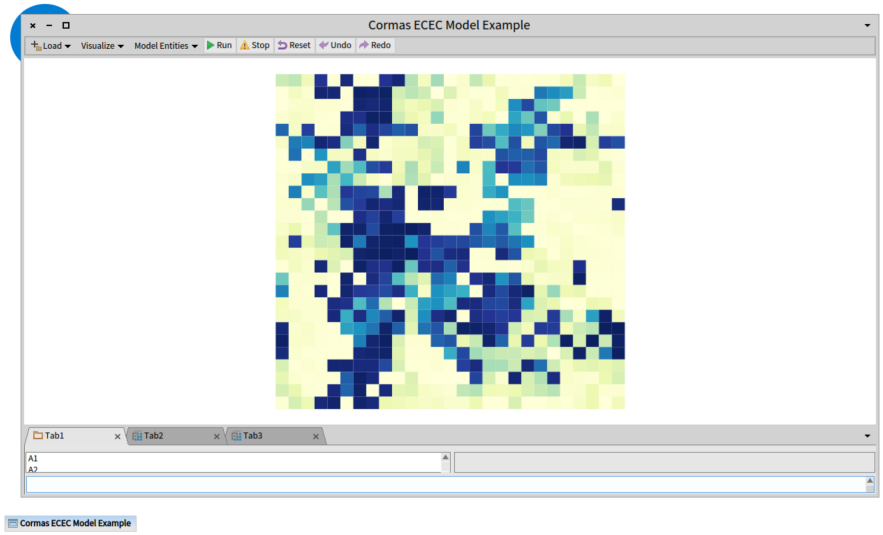


Figure 3. CORMAS ECEC Model simulation on the run.

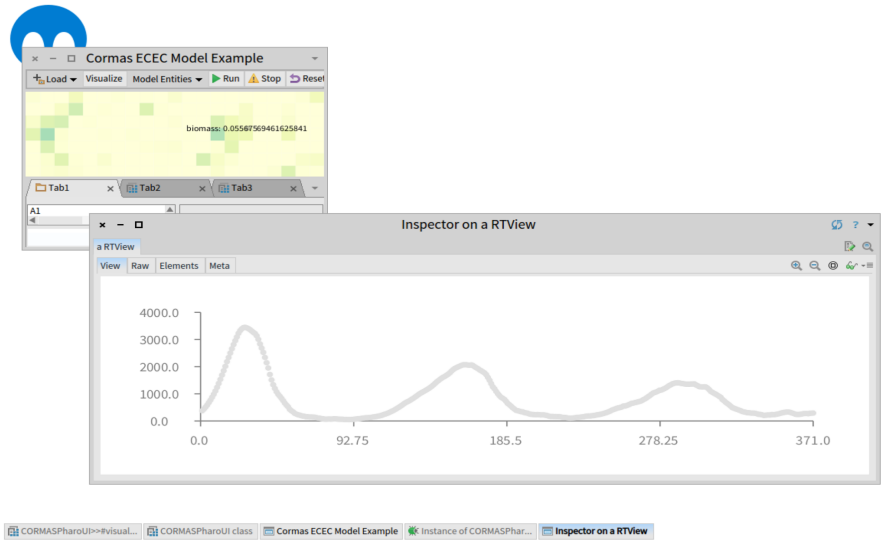


Figure 4. Simple visualization of the data collected in the course of CORMAS ECEC Model simulation

4. Conclusion and Future Work

Our brief analysis of the four well-known modelling and simulation platforms namely MobiDyC, NetLogo, GAMA and CORMAS highlights the specificities of each one, while presenting their advantages and disadvantages concisely spanning programming flexibility, portability, extensibility, scalability and heterogeneous data support. In addition, it spells out the need to explore lively simulations to feed the necessity of interaction with the agents in a simulation effortlessly in real-time. It is in this context that we have recently begun experimenting with porting CORMAS agent-modelling platform to the open-source Pharo platform. The rationale behind this choice is to allow a modern recreation of the platform that can take advantage of developments in: agile visualization engines, moldable tools, UI frameworks, DSLs and others.

In this work we focused our attention on the salient user interface component of ABM systems, namely the live spatial interface and presented our current progress in terms of supporting extensible visualizations and tools in case of CORMAS.

We observed that the new orientations of ABM platforms require development of extensions and plug-ins compatible with third-party applications for better interoperability. It turns out that the world of open-source developers enables by its dynamism to obtain reliable and quickly available extensions. This is the case notably with Pharo, a version of which already runs on many platforms including Android, whereas this specificity is not yet considered by Cincom-VW.

The migration of CORMAS from VW to Pharo was recently initiated, but it already offers considerable prospects because it exhibits an in-progress version which is well within the reach of

RIGHTSLINK

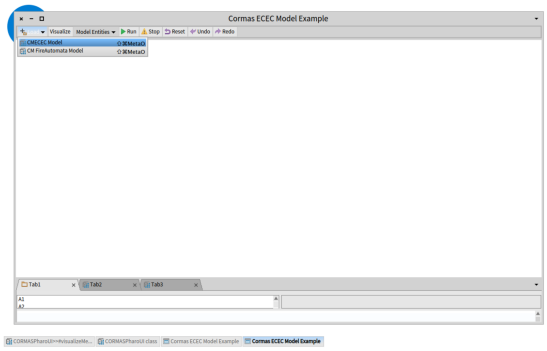


Figure 5. Loading different models viz. CM ECEC Model and the Fire Automata Model in the latest Pharo extension of CORMAS.

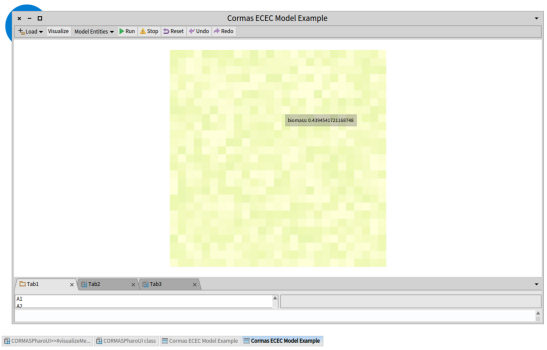


Figure 6. Presenting cell information (here: *biomass*) using customized inspector initiated on *onHover* event.

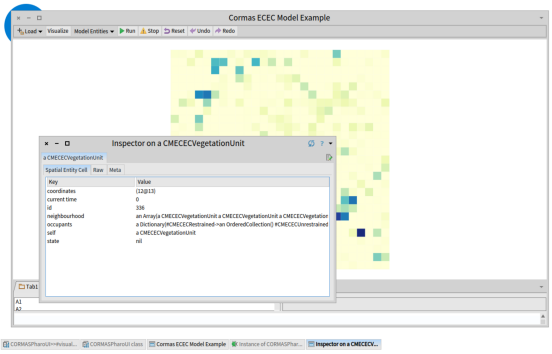


Figure 7. Presenting detailed cell information (here: *id*, *coordinates*, etc.) using customized inspector initiated on *onClick* event.

the current official proprietary version by simply plying existing UI frameworks and the visualization engines in Pharo.

The complementary and future work in this area is now to improve the graphical user interface (zoom and multi-windowing), to include all the missing features (such as time traveling options) and to design better techniques for data analysis and visualization. To help domain experts to design and implement their models by reducing their dependency on computer scientists for coding, the future developments of CORMAS are focused on the design of embedded DSLs and user interfaces to automatically generate a part of the computer code. Towards this direction, a class diagram editor that generates the class structure and their attributes and translates the associations into code [34] is already available and will be integrated in the new CORMAS framework. The intention is also to extend CORMAS to study federated ABM systems as in the works of Casalicchio et al [35] who federated *Repast* (an agent-based simulation engine) and *OMNeT++* (an IT systems and communication networks modeling and simulation environment) to exploit simulation model reuse and expertise sharing in a multi-sectorial fields.

As the main orientation of CORMAS's future remains the interactivity with stakeholders and local actors, the major effort is dedicated on human-machine interfaces and ergonomic design. The migration of CORMAS under Pharo allows us to make possible the development of custom-made IT tools for supporting the collective design of fair and sustainable scenarios.

Acknowledgments

This work was funded by the FuturAgua Project (Belmont Forum) with a complementary scholarship from European Smalltalk Users Group (ESUG) at IRD (Institut de recherche pour le développement) of Bondy, France. Due thanks to Pierre Bommel (CIRAD), Serge Stinckwich (IRD), Nick Papoylias (IRD) and Stéphane Ducasse (INRIA) for their continuous support and guidance.

References

[1] A. C. Kay, "The Early History of Smalltalk," *SIGPLAN Not.*, 1993.

[2] A. Goldberg and D. Robson, "Smalltalk-80: The Language and its Implementation," *Xerox Palo Alto Research Center*, 1983.

[3] L. Gustafsson and M. Sternad, "Consistent micro, macro, and state-based population modelling," *Mathematical Biosciences*, 2010.

[4] V. Grimm and S. F. Railsback, "Individual-based Modeling and Ecology," *Princeton Series in Theoretical and Computational Biology*, 2005.

[5] S. Tisue and U. Wilensky, "NetLogo: A simple environment for modeling complexity," *International Conference on Complex Systems, Boston*, 2004.

[6] J. W. Pepper and B. B. Smuts, "The evolution of cooperation in an ecological context: An agent-based model," *Dynamics of human and primate societies: agent-based modeling of social and spatial processes*, 2000.

[7] C. Nikolai and G. Madey, "Tools of the Trade: A Survey of Various Agent Based Modeling Platforms," *Journal of Artificial Societies and Social Simulation vol. 12, no. 2 2*, 2009.

[8] C. M. Macal and M. J. North, "Tutorial on agent-based modelling and simulation," *Proceedings of the Winter Simulation Conference*, 2005.

[9] V. Ginot and C. Le Page, "Mobidyc, a generic multi-agents simulator for modeling populations dynamic," *Proceedings of the 11th International Conference on Industrial and Engineering Applications of Artificial In Telligence and Expert Systems: Tasks and Methods in Applied Artificial Intelligence*, 1998.

[10] V. Ginot, "MobiDyC: Summary," *The Institut national de la recherche agronomique (INRA)*, 2008.

[11] S. Souissia, L. Seuronta, F. G. Schmitta, and V. Ginot, "Describing space-time patterns in aquatic ecology using IBMs and scaling and multi-scaling approaches," *Nonlinear Analysis: Real World Applications*, 2005.

[12] B. Axelsson, "Mindeditor: An End-User Tool for Implicit Design of Simulated Human Behaviour," *UMEÅ University, Department of Computing Science*, 2003.

[13] R. Allan, "Survey of Agent Based Modelling and Simulation Tools," *Science and Technology Facilities Council, Daresbury Laboratory Technical Reports*, 2010.

RIGHTSLINK

[14] S. Tisue and U. Wilensky, "Breeding faster turtles: Progress towards a Netlogo compiler," *Conference on Social Dynamics: Interaction, Reflexivity and Emergence, Chicago, IL*, 2006.

[15] S. Tisue and U. Wilensky, "Netlogo: Design and implementation of a multi-agent modeling environment," *Conference on Social Dynamics: Interaction, Reflexivity and Emergence, Chicago, IL*, 2004.

[16] U. Wilensky, "NetLogo User Manual - Interface Guide," *Northwestern University - Center for Connected Learning and Computer-Based Modeling (CCL)*, 2017.

[17] P. Taillandier, B. Gaudou, D.-A. Vo, A. Grignard, and A. Drogoul, "A Practical Introduction to the GAMA Agent-based Modeling Platform," *CS-DC'15 World e-conference*, 2015.

[18] N. R. Ramli, S. Razali, and M. Osman, "An Overview of Simulation Software for Non-Experts to Perform Multi-Robot Experiments," *The International Symposium on Agents, Multi-Agent Systems and Robotics*, 2015.

[19] B. Davies, "Turtle Power: An Ode to NetLogo," *Simulating Complexity*, 2014.

[20] "GAMA, modeling made easy," *IRD/UPMC international research unit UMMISCO, MSI Research Team - Vietnam National University, CNRS/University of Rouen, CNRS/University of Toulouse 1, DREAM Research Team - University of Can Tho and CNRS/University Paris-Sud*, 2016.

[21] F. Bousquet, I. Bakam, H. Proton, and C. L. Page, "CORMAS: Common-Pool Resources and Multi-agent Systems," *Proceedings of the 11th International Conference on Industrial and Engineering Applications of Artificial In Telligence and Expert Systems: Tasks and Methods in Applied Artificial Intelligence*, 1998.

[22] P. Bommel, N. Becu, C. L. Page, and F. Bousquet, "Participatory Agent-based Simulation for Renewable Resource Management: The Role of the CORMAS Simulation Platform to Nurture a Community of Practice," *Journal of Artificial Societies and Social Simulation*, 2012.

[23] P. Bommel, N. Becu, C. L. Page, and F. Bousquet, "CORMAS: An Agent-based Simulation Platform for Coupling Human Decisions with Computerized Dynamics," *Simulation and Gaming in the Network Society*, pp. 387–410, 2016.

[24] K. Bajracharya and R. Duboz, "Comparison of Three Agent-based Platforms on the Basis of a Simple Epidemiological Model (wip)," in *Proceedings of the Symposium on Theory of Modeling & Simulation - DEVS Integrative M&S Symposium*, DEVS 13, (San Diego, CA, USA), pp. 7:1–7:6, Society for Computer Simulation International, 2013.

[25] A. Bergel, B. Thi Mai, S. Stinckwich, and N. Papoulias, *Agile Visualization*, ch. Expressing Epidemiological Models. Lulu, 2016.

[26] A. Chiş, O. Nierstrasz, A. Syrel, and T. Gîrba, "GTInspector: a moldable domain-aware object inspector," in *Companion Proceedings of the 2015 ACM SIGPLAN International Conference on Systems, Programming, Languages and Applications: Software for Humanity*, pp. 15–16, ACM, 2015.

[27] L. Renggli, M. Denker, and O. Nierstrasz, "Language Boxes," in *International Conference on Software Language Engineering*, pp. 274–293, Springer, 2009.

[28] J. Brauer, "The Visualworks Development Environment, in Programming Smalltalk : Object-Orientation from the Beginning," *Springer Fachmedien Wiesbaden*, 2015.

[29] J. Fabry and S. Ducasse, "SpecUI Framework," *Spec Booklet - Pharo Org*, 2017.

[30] A. Chiș, T. Girba, J. Kubelka, S. Reichhart, and A. Syrel, "The Glamorous Toolkit," *http://gtoolkit.org*, 2017.

[31] L. Renggli, J. Kurs, and G. Larcheveque, "PetitParser: Building Modular Parsers," *Deep Into Pharo by Pharo Org*, 2017.

[32] A. Bergel, B. Thi Mai, S. Stinckwich, and N. Papoulias, *Agile Visualization*, ch. Domain-Specific Visualization Made Easy with Builders by Yuriy Tymchuk. Lulu, 2016.

[33] P. Bommel, N. Becu, C. L. Page, and F. Bousquet, "Building a CORMAS model from scratch step by step: The ECEC Model," *CIRAD, GREEN Research Unit*, 2012.

[34] P. Uhnak and P. Bommel, "Facilitating the Design of ABM and the Code Generation to Promote Participatory Modelling," *Proceedings of the 8th International Congress on Environmental Modelling and Software, July 10-14, Toulouse, FRANCE.*, 2016.

[35] E. Casalicchio, E. Galli, and S. Tucci, "Federated Agent-based Modeling and Simulation approach to study interdependencies in IT critical infrastructures," in *Distributed Simulation and Real-Time Applications, 2007. DS-RT 2007. 11th IEEE International Symposium*, pp. 182–189, Oct 2007.

RIGHTSLINK

Astrophys Space Sci (2020) 365:2
https://doi.org/10.1007/s10509-019-3713-y

ORIGINAL ARTICLE

Probing bulk viscous matter-dominated model in Brans-Dicke theory

C.P. Singh[1] · **Simran Kaur**[1]

Received: 17 September 2019 / Accepted: 18 December 2019
© Springer Nature B.V. 2019

Abstract We explore a matter-dominated model with bulk viscosity in Brans-Dicke theory to interpret the observed cosmic accelerating expansion phenomena. We obtain the exact solution of the field equations by taking constant bulk viscous coefficient (which potentially could explain the present accelerated expansion of the Universe) and Brans-Dicke scalar field proportional to some power of the scale factor. The model is studied statistically using the available astronomical data and then compare using the tools taken from information criterion. Using the best-fit values of model parameters we find that the model shows transition from decelerated phase to accelerated phase. The effective equation of state parameter is within the quintessence region ($-1 \leq \omega < 0$). Thus, the model shows quintessence behavior and approaches to ΛCDM model in late time. It does not cross the phantom divide line and hence it is free from big-rip singularity. The viscous model also alleviates the age problem. We discuss two diagnostic parameters, namely statefinder and $Om(z)$ and compare the model with the existing models by plotting the trajectories. We also apply Akaike information criterion (*AIC*) and Bayesian information criterion (*BIC*) to discriminate the viscous model based on the penalization associated to the number of parameters. The analyses based on the *AIC* indicates that there is less support for the viscous model when compared to the ΛCDM model, while those based on the *BIC* indicates that there is strong evidence against it in favor of the ΛCDM model.

✉ C.P. Singh
cpsphd@rediffmail.com

S. Kaur
skjuneja.maths@gmail.com

[1] Department of Applied Mathematics, Delhi Technological University (Formerly Delhi College of Engineering) Bawana Road, Delhi 110 042, India

Keywords Cosmology · FRW model · Brans-Dicke theory · Bulk viscosity

1 Introduction

The measurements of the luminosity-redshift relations observed for a number of newly discovered type Ia supernovae (Riess et al. 1998; Perlmutter et al. 1998, 1999) and cosmic microwave background radiation (CMBR) (Spergel et al. 2003, 2007) and large scale structure (Bennet et al. 2003; Tegmark et al. 2004) indicate that the Universe is expanding at present with an accelerated rate. Various models have been presented to explain this recent acceleration, one of them is the assumption of the existence of mysterious component so called *dark energy* (DE). It is assumed that DE has negative pressure which provides the dynamical mechanism for the accelerated expansion of the Universe. The simplest candidate of dark energy is the cosmological constant Λ, with equation of state $\omega_\Lambda = -1$ (Weinberg 1989). However, it suffers two well-known problems named "fine-tuning problem" and "cosmic coincidence problem" (Copeland et al. 2006). To solve these theoretical puzzles, various DE models, e.g., quintessence (Chiba et al. 2000; Caldwell 2002; Guo et al. 2005; Capozziello 2006; Zhang and Gui 2006; Feng 2008), Kessence (Scherrer 2004), phantom energy (Caldwell et al. 2003; Vikman 2005; Chimento and Richarte 2013), quintom (Zhao 2005; Wei and Cai 2008), exotic Chaplygin gas (Setare 2007), etc. have been proposed in the past few years; unfortunately, so far the nature of DE still remains a complete mystery.

In recent years it has also been attempted in modelling the missing energy of the universe and to explain its late time accelerated expansion in view of scalar tensor theories. The

Published online: 31 December 2019

scalar-tensor theories of gravity are some of the most established and well studied alternative theories of gravity that exist in the literature. The interest in scalar-tensor theories of gravity arises from inflationary cosmology, supergravity and superstring theory. The pioneering study on scalar-tensor theories was carried out by Brans and Dicke (1961), known as Brans-Dicke (BD) theory. This theory modifies the Einstein's gravity in a way admitting the so-called "Mach's principle". A new scalar degree of freedom is added to incorporate the Mach's principle. This was the first gravity theory in which the dynamics of gravity were described by a scalar field, while space–time dynamics were represented by the metric tensor. This scalar field has a very active role for the description of the early era as well as the present phase of the evolution of the Universe. This theory also passes the observational tests in the solar system domain (Bertotti et al. 2003) and provides an explanation of the accelerated expansion of the Universe (Das and Banerjee 2008). In BD theory the gravitational constant G varies as the inverse of the scalar field ϕ, i.e., $G \sim \frac{1}{\phi(t)}$, which couples to gravity with a coupling parameter ω (for review, see Ref. Clifton et al. 2012). The BD theory provides useful clues to the solutions for some of the outstanding problems in cosmology and could generate sufficient acceleration in the matter dominated Universe.

The concept of viscosity in cosmology has a long history but their application to the accelerating universe are rather recent. The viscous mechanism has gained immense interest in recent times for various reasons. Various aspects of the theory of viscous cosmology have been discussed in Refs. Grøn 1990; Brevik and Odintsov 2002; Brevik and Hallanger 2004; Singh 2008; Nojiri and Odintsov 2007; Brevik et al. 2010; Velten et al. 2013; Wang and Meng 2014; Bamba and Odintsov 2016; Brevik et al. 2017. Eckart (1940) was first to propose the general theory of dissipation in relativistic imperfect fluid, and later on modified by Landau and Lifshitz (1987). Israel and Stewart (1976) developed a relativistic second order full causal theory. For review of non-causal and causal theories, we refer to Refs. Grøn 1990; Maartens 1995; Brevik and Grøn 2013. The idea of viscous dark energy models has been presented in different ways to understand the evolution of the universe. Ren and Meng (2006) investigated the universe evolution through a cosmological model with bulk viscosity. Tawfik and Harko (2012) discussed phase transition of the viscous early universe. Singh and Kumar (2014, 2016) explored different aspects of viscosity in $f(R,T)$ gravity taking Hubble horizon as IR cut-off. Viscous cosmology with inhomogeneous equation of state have also been proposed and studied in Refs. Nojiri and Odintsov 2005; Capozziiello et al. 2006.

One of the most intriguing questions concerning the late time accelerating expansion in the Universe is that, the possibility of distinguishing between cosmological constant and dark energy models. High sensitivity as well as model independent properties should be considered for introducing a reliable diagnostic measure. Using diagnostic analysis, we are able to study some kinematic properties of the viscous dark energy model. In this context, Sahni et al. (2003) introduced a new geometrical diagnostic pair $\{r,s\}$, known as statefinder parameters, which is constructed from the scale factor and its derivatives up to the third order. The statefinder pair $\{r,s\}$ is geometrical in the nature as it is constructed from the space–time metric directly. Therefore, the statefinder parameters are more universal parameters to study the DE models than any other physical parameters. In a flat ΛCDM model, the statefinder pair has a fixed point value $\{r,s\}=\{1,0\}$ where as it is $\{r,s\}=\{1,1\}$ for SCDM model. One can plot the trajectories in $r-s$ and $r-q$ planes to discriminate various DE models. We will discuss the statefinder diagnostic for viscous model and obtain the fixed point values of statefinder pair under suitable constraints.

As a complementary to $\{r,s\}$, Sahni et al. (2008) proposed another new diagnostic called, $Om(z)$. This diagnostic helps to distinguish the present matter density Ω_{om} in different models more effectively. The Om(z) diagnostic is indeed a geometrical diagnostic which combines Hubble parameter and redshift. It can differentiate dark energy model from ΛCDM. For ΛCDM model $Om(z)=\Omega_{0m}$ while for other dark energy models, $Om(z)$ depends on redshift. Phantom like dark energy corresponds to the positive slope of $Om(z)$ whereas the negative slope means dark energy behaves like quintessence.

The purpose of this paper is to explore the role of bulk viscosity in BD theory. Such a viscous term acts as a negative pressure. It is well known that bulk viscosity can generate early and late time accelerated expansion. But, such possibility has been investigated only in the context of the primordial Universe, concerning also the search of non singular models (Murphy 1973; Pimentel 1985). Indeed, it has been shown (Fabris et al. 2006) that for the right viscosity coefficient, an accelerating cosmology can be achieved without the need of a cosmological constant. We will consider a simple bulk viscosity model, in the context of the Eckart formalism (Eckart 1940). In spite of the problems of this theory, it has been used to model bulk viscous model as responsible for the observed acceleration of the Universe assuming that the approximation of vanishing relaxation time is valid (Avelino and Nucamendi 2009). Avelino and Nucamendi (2008a, 2008b) studied and constrained a matter-dominated model with bulk viscosity using observational data and found that the model is viable just if the bulk viscosity is triggered in recent times in case bulk viscous coefficient is proportional to the Hubble parameter. However, the model is ruled out by the observations when bulk viscous coefficient is taken to be constant. In this paper, we extend this work to matter-dominated model with constant

bulk viscous coefficient in scalar-tensor theory proposed by Brans and Dicke (1961) and constrain the model to observe the cosmic acceleration.

The paper is organized as follows: In Sect. 2 we consider the matter-dominated bulk viscous model in the framework of Friedmann-Roberston-Walker (FRW) line element in BD theory. The solutions of field equations are presented in Sect. 3 and discuss the evolution of the scale factor and deceleration parameter. In Sect. 4 we discuss two diagnostic parameters, namely statefinder parameter and *Om* parameter to discriminate our model with ΛCDM model. Section 5 presents the observational data of Type Ia Supernova, Observational Hubble data and Baryon Acoustic Oscillation data to constrain our viscous model parameters. In Sect. 6 we analyze the constrain of model parameters in the solutions and accordingly discuss the evolution of our model. Section 7 deals the information criteria to select the model. Finally, in Sect. 8 we conclude the outcomes of our work.

2 Viscous model in Brans-Dicke theory

In the context of the Brans-Dicke theory (1961), the action in the presence of matter with Lagrangian $\mathcal{L}_m$ in Jordan frame is given by Kim (2005)

$$S=\int d^4x\sqrt{-g}\left[\frac{1}{16\pi}\left(\phi R-\omega\frac{\nabla_\alpha\phi\nabla^\alpha\phi}{\phi}\right)+\mathcal{L}_m\right], \tag{1}$$

where ϕ is the BD scalar field representing the inverse of Newton's constant which is allowed to vary with space and time, ω is the generic dimensionless coupling parameter of the theory and $\mathcal{L}_m$ is the Lagrangian density for all matter fields excluding the scalar field ϕ. The other symbols have their usual meanings.

The gravitational field equations derived from the action (1) by varying the action with respect to the metric is

$$\begin{aligned}R_{\mu\nu}-\frac{1}{2}g_{\mu\nu}R=&\frac{8\pi}{\phi}T_{\mu\nu}\\&+\frac{\omega}{\phi^2}\left(\nabla_\mu\phi\nabla_\nu\phi-\frac{1}{2}g_{\mu\nu}\nabla_\alpha\phi\nabla^\alpha\phi\right)\\&+\frac{1}{\phi}\left(\nabla_\mu\nabla_\nu\phi-g_{\mu\nu}\nabla_\alpha\nabla^\alpha\phi\right),\end{aligned} \tag{2}$$

where $T_{\mu\nu}$ represents the energy-momentum tensor of the viscous matter field. We have assumed the matter content of the universe to be composed of a fluid represented by the energy momentum tensor

$$T_{\mu\nu}=(\rho_m+P)u_\mu u_\nu+P\,g_{\mu\nu} \tag{3}$$

where ρ_m and P are the energy density and effective pressure of the fluid, respectively, and u_μ is the four velocity of the fluid, i.e., $u_\mu u^\nu=-1$.

The wave equation that follows from Eq. (1), by varying the action with respect to the scalar field ϕ, is

$$\nabla_\alpha\nabla^\alpha\phi=\frac{8\pi}{(3+2\omega)}T^\lambda_\lambda, \tag{4}$$

where T^λ_λ is the trace of the energy-momentum tensor. Let us assume that the universe is homogeneous, isotropic and spatially flat. Such a universe is described by the Friedmann-Robertson-Walker (FRW) line-element

$$ds^2=-dt^2+a^2(t)\left[dr^2+r^2\left(d\theta^2+\sin^2\theta d\phi^2\right)\right], \tag{5}$$

where $a(t)$ is referred to as the scale factor and t is the cosmic time.

In an isotropic and homogeneous FRW universe, the dissipative effects arise due to the presence of bulk viscosity ζ in cosmic fluids. The bulk viscosity introduces dissipation by only redefining the effective pressure, P, according to

$$P=p_m+\Pi=p_m-3\zeta H, \tag{6}$$

where $\Pi=-3\zeta H$ is the bulk viscous pressure, ζ is the bulk viscous coefficient and H is the Hubble parameter.

The energy-momentum tensor (3) for bulk viscosity as the only dissipative phenomena modifies to

$$T_{\mu\nu}=\rho_m u_\mu u_\nu+(g_{\mu\nu}+u_\mu u_\nu)P. \tag{7}$$

The Eckart theory has problems in several aspects. One of them is that all the equilibrium states in this theory are unstable. Another one is that signals can propagate through the fluids with superluminal velocities. However, despite of these problems, it is widely accepted by several authors to model bulk viscosity as responsible for the recent acceleration by assuming that the approximation is valid for this purpose, i.e., assuming the vanishing relaxation time. In this limit the Eckart theory is a good approximation to the Israel-Stewart theory.

In such a background, we get the evolution equations of the Universe from (2):

$$H^2+H\frac{\dot\phi}{\phi}-\frac{\omega}{6}\frac{\dot\phi^2}{\phi^2}=\frac{8\pi}{3\phi}\rho_m, \tag{8}$$

$$2\frac{\ddot a}{a}+H^2+\frac{\omega}{2}\frac{\dot\phi^2}{\phi^2}+2H\frac{\dot\phi}{\phi}+\frac{\ddot\phi}{\phi}=-\frac{8\pi}{\phi}P \tag{9}$$

where $H=\frac{\dot a}{a}$ is the Hubble parameter and the dot denotes derivative with respect to cosmic time t. The scalar field evolution equation is

$$\ddot\phi+3H\dot\phi=\frac{8\pi}{2\omega+3}(\rho_m-3P). \tag{10}$$

The wave equation (10) combines with (8) and (9) lead to the energy conservation equation

$$\dot\rho_m+3(\rho_m+P)H=0. \tag{11}$$

We work in the Jordan frame. One interesting thing about BD theory in the Jordan frame is that the conservation equation holds for the matter and scalar field separately. Or in a slightly different way, the Bianchi identity along with the wave Eq. (4) gives the matter conservation equation

Assuming that at the present epoch the universe is filled with cold matter of negligible pressure (i.e., dust), we put $p_m = 0$ in the last equation. So, P will have a simple form of $-3\zeta H$. We can principally solve the matter density ρ_m with respect to cosmic time t when the explicit form of ζ is specified. The conservation equation (11) now becomes

$$\dot{\rho_m} + 3(\rho_m - 3\zeta H)H = 0, \tag{12}$$

which can be written in terms of scale factor as

$$a\frac{d\rho_m}{da} = 3(3\zeta H - \rho_m). \tag{13}$$

Equation (13) is valid for any parametrization of ζ. However, in this paper, we will work with the simplest parametrization for the bulk viscous coefficient $\zeta = const. = \zeta_0$. We assume the bulk viscous coefficient to be positive, i.e., $\zeta_0 > 0$ to preserve the validity of second law of thermodynamics. We shall assume that the Brans-Dicke scalar field can be described as a power $-$ law of the scale factor, $\phi = \phi_0\, a^n$, where ϕ_0 and n are constants (Banerjee and Pavon 2007; Sheykhi 2010). Equation (8) can be rewritten as

$$H^2 = \frac{16\pi}{(6 + 6n - \omega n^2)\phi_0}\frac{\rho_m}{a^n}. \tag{14}$$

It is to be noted that the solar system experiments Cassini predict a very stringent high bound result on ω as $|\omega| > 40000$ (Bertotti et al. 2003; Will 2006). However, the cosmological observations put relatively lower bounds on ω (Wu and Chen 2010).

Substituting (14) into (13) we obtain

$$a\frac{d\rho_m}{da} + 3\rho_m - \gamma\frac{\rho_m^{1/2}}{a^{n/2}} = 0, \tag{15}$$

where $\gamma = 9(\frac{16\pi}{(6+6n-\omega n^2)\phi_0})^{1/2}\,\zeta_0$.

On solving (15), we get the evolution for energy density as

$$\rho_m = \left[\frac{\gamma}{(3-n)}a^{-n/2} + \left(\rho_{m0}^{1/2} - \frac{\gamma}{(3-n)}\right)a^{-3/2}\right]^2 \tag{16}$$

where ρ_{m0} is the present value of energy density. We assume that the value of scale factor at present is $a_0 = 1$. Now, by considering that $a = (1+z)^{-1}$, the energy density can be rewritten as

$$\rho_m = \left[\frac{\gamma}{(3-n)}(1+z)^{n/2} + \left(\rho_{m0}^{1/2} - \frac{\gamma}{(3-n)}\right)(1+z)^{3/2}\right]^2 \tag{17}$$

Inserting (17) into (14), we obtain

$$H(z) = H_0(1+z)^{-\frac{n}{2}} \times\left[\frac{\bar{\zeta_0}(1+z)^{\frac{n}{2}}}{(3-n)} + \left(\Omega_{m0}^{1/2} - \frac{\bar{\zeta_0}}{(3-n)}\right)(1+z)^{\frac{3}{2}}\right] \tag{18}$$

where H_0 is the present Hubble constant, $\gamma = \bar{\zeta_0}/H_0$ is the dimensionless bulk viscous coefficient and $\Omega_{m0} = \rho_{m0}/\rho_{crit}^0$, where $\rho_{crit}^0 = 3\phi_0\, H_0^2/8\pi$ is the present critical density. Taking $\Omega_{m0} = 1$ at present as bulk viscous matter is the only component, Eq. (18) finally becomes

$$H(z) = \frac{H_0}{(3-n)}(1+z)^{-\frac{n}{2}} \times\left[\bar{\zeta_0}(1+z)^{\frac{n}{2}} + (3-n-\bar{\zeta_0})(1+z)^{\frac{3}{2}}\right] \tag{19}$$

It should be noticed that for $n = 0$, the above expression reduces to the solution derived in Ref. (Avelino and Nucamendi 2009).

3 Solution of viscous model

Equation (19) has the solution

$$\int_1^a \frac{\acute{a}^{\frac{3}{2}-\frac{n}{2}}\, d\acute{a}}{\acute{a}[\bar{\zeta_0}\acute{a}^{\frac{3}{2}-\frac{n}{2}} + (3-n-\bar{\zeta_0})]} = \frac{H_0}{(3-n)}\int_{t_0}^t d\acute{t} \tag{20}$$

where t is the cosmic time and t_0 is the present cosmic time. In the absence of bulk viscous term, $\bar{\zeta_0} = 0$, equation (20) gives the scale factor for matter-dominated universe in BD theory as

$$a(t) = \left(\frac{3-n}{2}H_0(t-t_0) + 1\right)^{\frac{2}{3-n}}, \quad (3-n) \neq 0, \tag{21}$$

which is power-law of expansion. The model decelerates for $0 \leq n < 1$ and accelerates for $n \geq 1$.

For $\bar{\zeta_0} \neq 0$, Equation (20) gives

$$a = \left[\frac{(3-n)\exp(\frac{\bar{\zeta_0}H_0}{2}(t-t_0)) - (3-n-\bar{\zeta_0})}{\bar{\zeta_0}}\right]^{\frac{2}{(3-n)}}, \tag{22}$$

where $\bar{\zeta_0} \neq 0$ and $(3-n) \neq 0$.

The transition from one phase to another phase can also be obtained by defining the deceleration parameter q, which is defined as $q = -a\ddot{a}/\dot{a}^2$. It is straightforward to show from (22) that the deceleration parameter takes the following form

$$q(t) = -\frac{1}{2}\left[2 - (3-n-\bar{\zeta_0})e^{-\frac{\bar{\zeta_0}H_0}{2}(t-t_0)}\right]. \tag{23}$$

In terms of redshift, q is given by

$$q(z,\bar{\zeta_0}) = -1 + \frac{1}{2}\frac{(3-n-\bar{\zeta_0})(3-n)}{\bar{\zeta_0}\,(1+z)^{-\frac{(3-n)}{2}} + (3-n-\bar{\zeta_0})} \tag{24}$$

We find that the value of q is time-dependent and hence the model shows the transition. We see that $q = \frac{(1-n)}{2}$ in the absence of bulk viscous term, i.e., for $\bar{\zeta_0} = 0$. Thus, for $0 \leq n < 1$, the model decelerates and accelerates for $n \geq 1$.

The solution of BD scalar field is given by

$$\phi = \phi_0\left[\frac{(3-n)\exp(\frac{\bar{\zeta_0}H_0}{2}(t-t_0)) - (3-n-\bar{\zeta_0})}{\bar{\zeta_0}}\right]^{\frac{2n}{(3-n)}}. \tag{25}$$

Now, we will discuss the solutions for three different values of $\bar{\zeta_0}$ which are as follows:

Case (i): $0 < \bar{\zeta_0} < (3-n)$

In this case, we observe that as $t \to \infty$, the scale factor tends to have the form like de Sitter Universe, $a \sim \exp(\frac{\bar{\zeta_0}H_0}{(3-n)}(t-t_0))$. It is obvious that for any value of $\bar{\zeta_0}$ in $0 < \bar{\zeta_0} < (3-n)$, the model predicts a Big-bang in past during time $t_B = t_0 + \frac{2}{\bar{\zeta_0}H_0}\ln(1 - \frac{\bar{\zeta_0}}{(3-n)})$. The age of the Universe, the elapsed time between Big-bang time t_B until today t_0 is given by $|t_B - t_0| = \frac{2}{\bar{\zeta_0}H_0}\ln(1 - \frac{\bar{\zeta_0}}{(3-n)})$. Thus, there is decelerating epoch followed by a transition from deceleration to acceleration epoch in late time. To compute the value of scale factor where transition happens, we have from (19)

$$\frac{d\dot{a}}{da} = \frac{H_0}{(3-n)}\left[\bar{\zeta_0} + \frac{(n-1)(3-n-\bar{\zeta_0})}{2}\,a^{\frac{n-3}{2}}\right]. \tag{26}$$

Equating (26) to zero, we get

$$a_T = \left[\frac{(1-n)(3-n-\bar{\zeta_0})}{2\bar{\zeta_0}}\right]^{\frac{2}{3-n}}, \tag{27}$$

where the subscript "T" denotes the transition.

We plot the evolution of the scale factor as a function of time in Fig. 1 for different values of n and $\bar{\zeta_0}$ in $0 \leq \bar{\zeta_0} \leq (3-n)$. It is observed that the transition from decelerated phase to accelerated one occurs in future ($a_T > 1$) for the value of $\bar{\zeta_0}$ in interval $0 < \bar{\zeta_0} < (1-n)$. When $\bar{\zeta_0} = (1-n)$, we get $a_T = 1$, i.e., the transition takes place today whereas the transition occurs in past ($0 < a_T < 1$) for the values of $\bar{\zeta_0}$ in $(1-n) < \bar{\zeta_0} < (3-n)$. We observe that as the value of $\bar{\zeta_0}$ increases, the transition from deceleration to acceleration occurs in the past. In the limiting case when $\bar{\zeta_0}$ approaches to $(3-n)$, $a_T \to 0$, i.e., the transition takes place very close to Big-bang.

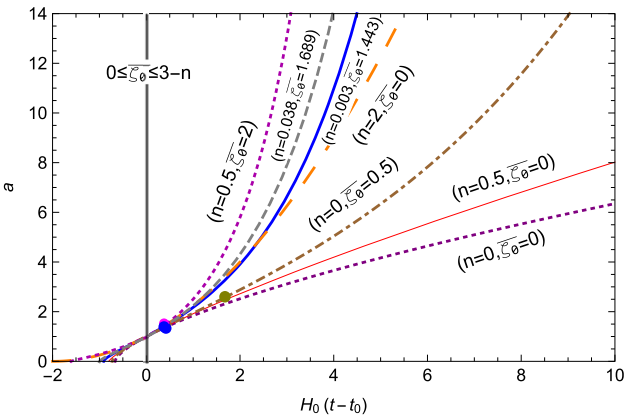


Fig. 1 The scale function as a function of time for different values of n and $\bar{\zeta_0}$ in the range of $0 \leq \bar{\zeta_0} \leq (n-3)$. The grey dotted line represents the evolution coming from *SNe* + *OHD* and the blue solid line corresponds trajectory coming from *SNe* + *OHD* + *BAO* data set analysis (see, Sect. 6). The dots show the transition point

In this case, the deceleration parameter decreases from $q(1) = -1 + \frac{1}{2}(3-n-\bar{\zeta_0})$ to $q(\infty) = -1$ which shows the transition from positive to negative. When $\bar{\zeta_0} = (1-n)$, $q = 0$, i.e., the transition from decelerated to accelerated phase takes place today.

Case (ii): $\bar{\zeta_0} = (3-n)$

In this case, Eq. (21) gives $a(t) = \exp(H_0(t-t_0))$ and $q = -1$ that corresponds to the de Sitter universe. The model has singularity at infinite past, i.e., as $(t-t_0) \to -\infty$, $a \to 0$ and as $(t-t_0) \to \infty$, $a \to \infty$. The matter density is constant, $\rho_m = \frac{(6+6n-\omega n^2)\phi_0 H_0^2}{16\pi(3-n)}$. In Fig. 1, the dotted line $n = 0.5$ and $\bar{\zeta_0} = 2.5$ corresponds to $\bar{\zeta_0} = (3-n)$ (the de Sitter universe).

Case (iii): $\bar{\zeta_0} > (3-n)$

In this case, we observe that the viscous model always accelerates. There is not any deceleration-acceleration transition. The deceleration parameter always gives negative value. In this case, the model starts with minimum scale factor (Einstein static universe) and behaves as a de Sitter universe in future. The evolution of the scale factor is shown in Fig. 2 for different values of n and $\bar{\zeta_0}$ in $\bar{\zeta_0} \geq (3-n)$.

4 Diagnostic analysis

In this section we discuss about the statefinder parameter $\{r, s\}$ (Sahni et al. 2003), which discriminates the proposed model with ΛCDM model. The diagnostic statefinder parameter $\{r, s\}$ is defined as:

$$r = \frac{\dddot{a}}{aH^3} \quad \text{and} \quad s = \frac{r-1}{3(q-1/2)}. \tag{28}$$

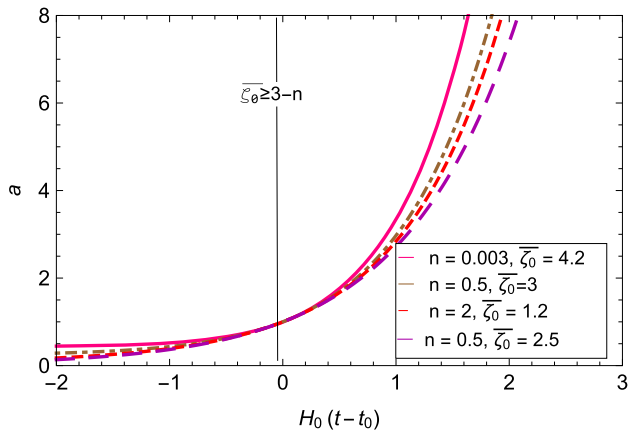


Fig. 2 Plot of the scale factor as function of time for different values of $\bar{\zeta}_0$ in the range of $\bar{\zeta}_0 \geq (n-3)$

For this viscous BD model, the statefinder parameters are given by

$$r = \frac{1}{4}e^{-\bar{\zeta}_0 H_0(t-t_0)}\left[(3-n)e^{\frac{\bar{\zeta}_0 H_0}{2}(t-t_0)} - (3-n-\bar{\zeta}_0)\right]^2 + \frac{3}{4}(n-1)e^{-\frac{\bar{\zeta}_0 H_0}{2}(t-t_0)} \times\left[(3-n)e^{\frac{\bar{\zeta}_0 H_0}{2}(t-t_0)} - (3-n-\bar{\zeta}_0)\right] + \frac{1}{2}(n-1)(n-2) \tag{29}$$

and

$$s = \frac{(r-1)}{3[-\frac{3}{2} + \frac{(3-n-\bar{\zeta}_0)}{2}e^{-\frac{\bar{\zeta}_0 H_0}{2}(t-t_0)}]} \tag{30}$$

We observe that as $(t-t_0) \to \infty$, the statefinder pair $\{r,s\} \to \{\frac{(n-1)(n-2)}{2}, \frac{2-(n-1)(n-2)}{9}\}$, which deviates from the ΛCDM model. However, it corresponds to the ΛCDM model for $n=0$.

Now, we discus another diagnostic parameter, known as $Om(z)$, which is defined as (Sahni et al. 2008)

$$Om(z) = \frac{\frac{H^2(z)}{H_0^2} - 1}{(1+z)^3 - 1}, \tag{31}$$

where H_0 is the present value of the Hubble parameter. It has been shown that the slope of Om can distinguish dynamical DE from the cosmological constant in a robust manner. A constant behavior of Om with respect to z signifies that DE is a cosmological constant. The positive slopes of $Om(z)$ give the phantom behavior whereas the negative slopes mean for quintessence.

Using (19) into (31), we get

$$Om(z) = \frac{[\bar{\zeta}_0 + (3-n-\bar{\zeta}_0)(1+z)^{\frac{(3-n)}{2}}] - (3-n)^2}{(3-n)^2[(1+z)^3 - 1]}. \tag{32}$$

In the case of $z=-1$, $Om(z) = 1 - \frac{\bar{\zeta}_0}{(3-n)^2}$, which is constant, which deviates from the ΛCDM model. However, in the absence of bulk viscosity we get $Om(z)=1$. We will discuss in details the trajectories of $\{r,s\}$ and $Om(z)$ in Sect. 6.

5 Observational data

In this section, we will constrain the viscous model by using the latest observational data, namely Supernovae Type Ia (SNe), observational Hubble parameter data (OHD) and Byronic Acoustic Oscillation (BAO) data. We compute the best estimated values of free parameters $(n, \bar{\zeta}_0, H_0)$ and confidence interval by χ^2 function. We use the minimized chi-square test to obtain the best-fitted value of the free parameters and likelihoods by using Markov Chain Monte Carlo (MCMC) method (Foreman-Mackey et al. 2012).

We use the cJLA data set of 31 check points (30 bins) covering the redshift range $z=[0.01, 1.3]$ (Betoule et al. 2014). Type Ia supernovae are standard candles and are useful in determining the expansion history of the universe. The supernova's apparent brightness determines its distance from the observer and the time taken by photons to reach the observer. By comparing the flux to redshift relation for numerous *SNe*, the rate of expansion of Universe and its variation with time can be determined. The relation of luminosity distance $d_L(z)$ to redshift z in a spatially flat Universe is given by

$$d_L(z,\theta) = c(1+z)\int_0^z \frac{dz'}{H(z',\theta)} \tag{33}$$

where $H(z)$ is defined in (19), θ denotes the model parameters, namely $\bar{\zeta}_0$, n, H_0 and c is the speed of light given in units of km/sec. The theoretical distance modulus is defined as

$$r = \mu_b - M - 5\,\log_{10}\,d_L, \tag{34}$$

where μ_b is the observational distance modulus and M is a free normalization parameter. We construct the statistical χ^2 function as

$$\chi^2_{SNe} = r^t C_b^{-1} r, \tag{35}$$

where C_b is the covariance matrix of μ_b, (see Table F.2 in Betoule et al. 2014). We numerically minimize χ^2 to compute the *best estimates* for free parameters $(\bar{\zeta}_0, n, H_0)$ of the model.

Next, we consider the cosmological observational data from *OHD*. We use the OHD dataset of 43 measurement points collected in Cao et al. (2018) in the redshift range

$0 < z < 2.5$. The χ^2 for Observational Hubble Data is

$$\chi^2_{OHD} = \sum_{i=1}^{m} \frac{[H(z_i) - H_{obs}(z_i,\theta)]^2}{\sigma_i^2}, \tag{36}$$

where $H(z_i)$ and $H_{obs}(z_i)$ are the theoretical and observed values respectively and σ_i^2 the standard deviation of each $H_{obs}(z_i)$.

We also include the sample of *BAO* distances measurements from *SDSS*(R) (Padmanabhan et al. 2012), the $6dF$ Galaxy survey (Beutler et al. 2011), *BOSS CMASS* (Anderson et al. 2014) and three parallel measurements from *WiggleZ* survey (Blake et al. 2012).

The angular diameter, $d_A(z,\theta)$ is given by

$$d_A(z_*,\theta) = c\int_0^{z_*} \frac{dz'}{H(z',\theta)} \tag{37}$$

where z_* denotes the photons decoupling redshift and according to the Planck 2015 results (Ade et al. 2016), its value is $z_* \approx 1090$. Further, the dilation scale, $D_v(z,\theta)$ is given by $D_v(z,\theta) = (\frac{d_A^2(z,\theta)cz}{H(z,\theta)})^{\frac{1}{3}}$.

The corresponding χ^2 function is given by Giostri et al. (2012)

$$\chi^2_{BAO} = A^T C^{-1} A \tag{38}$$

where A is a matrix given by

$$A = \begin{bmatrix} \frac{d_A(z_*,\theta)}{D_v(0.106,\theta)} - 30.84 \\ \frac{d_A(z_*,\theta)}{D_v(0.35,\theta)} - 10.33 \\ \frac{d_A(z_*,\theta)}{D_v(0.57,\theta)} - 6.72 \\ \frac{d_A(z_*,\theta)}{D_v(0.44,\theta)} - 8.41 \\ \frac{d_A(z_*,\theta)}{D_v(0.6,\theta)} - 6.66 \\ \frac{d_A(z_*,\theta)}{D_v(0.73,\theta)} - 5.43 \end{bmatrix}$$

and C^{-1} is the inverse of covariance matrix (Giostri et al. 2012). Here, we have adopted the correlation coefficients given in Ref. (Hinshaw et al. 2013).

Therefore, the total χ^2 is given by

$$\chi^2 = \chi^2_{SNe} + \chi^2_{OHD} + \chi^2_{BAO}, \tag{39}$$

where χ^2_{SNe} is given by (35), χ^2_{OHD} and χ^2_{BAO} are respectively given by (36) and (38). The best-fitted model parameters are determined by minimizing the total χ^2.

6 Results and analysis

In this section we perform best-fitting to the viscous model solutions obtained in Sects. 3 and 4 and plot the evolutions for cosmological and diagnostic parameters. We fit the viscous model to joint observational data of *SNe*, *OHD* and *BAO*. We use the MCMC method and finally we obtain the best $-$ fit parameters and the corresponding $\chi^2_{\min}$. The results of the estimation of the free parameters of the viscous BD model from the observational data sets, namely (i) $SNe + OHD$ and (ii) $SNe + OHD + BAO$, are presented in Table 1. The contour plots of parameters with $1\sigma(68.3\%)$ and $2\sigma(95.4\%)$ confidence level for $SNe + OHD$ and $SNe + OHD + BAO$ are shown in Figs. 3 and 4, respectively.

Table 1 The results of statistical analysis for the viscous BD model parameters and $\chi^2_{d.o.f.} = \chi^2_{\min}/(N-d)$, where N is the total data points and d is the number of free parameters

Parameters	SNe+OHD	SNe+OHD+BAO
$\bar{\zeta_0}$	$1.689^{+0.062}_{-0.087}$	$1.443^{+0.027}_{-0.030}$
n	$0.038^{+0.064}_{-0.025}$	$0.003^{+0.005}_{-0.002}$
H_0	$68.187^{0.566}_{-0.601}$	$65.324^{+0.487}_{-0.391}$
$\chi^2_{d.o.f.}$	0.493	0.931

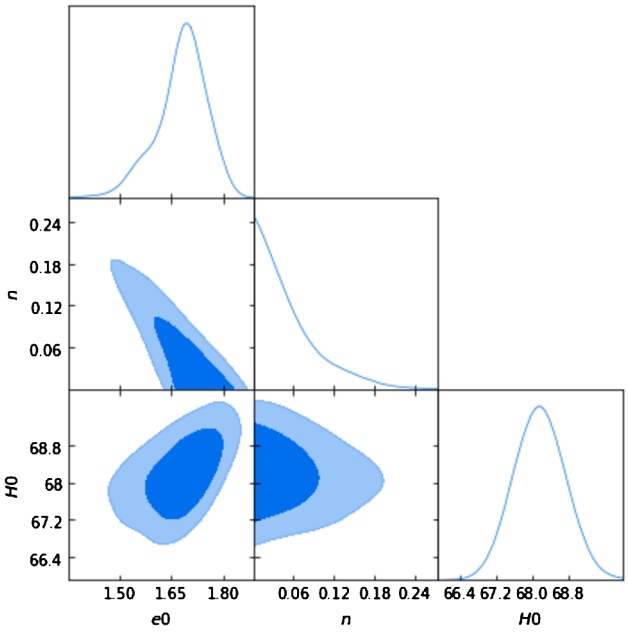


Fig. 3 The contour maps of $\bar{\zeta_0}$, n, H_0 at $1\sigma(68.3\%)$ and $2\sigma(95.4\%)$ confidence level using $SNe + OHD$. In Fig. $e0$ means $\bar{\zeta_0}$

The evolutions of the scale factor are shown in Figs. 1 and 2 for different values of model parameters. Using the best fitting data of parameters listed in Table 1, we plot the trajectories of the scale factor. In Fig. 1, the trajectories are shown for $0 < \bar{\zeta_0} < (3-n)$. The grey dotted line and the solid blue line represent the evolution of the scale factor for best fitting data of parameters obtained from $SNe + OHD$ and $SNe + OHD + BAO$, respectively. The trajectories show that the viscous universe decelerates in early time and accelerates in late time. Figure 2 shows the trajectories for $\bar{\zeta_0} \geq (3-n)$. The transition scale factor a_T for best fitting data, where the model transits from decel-

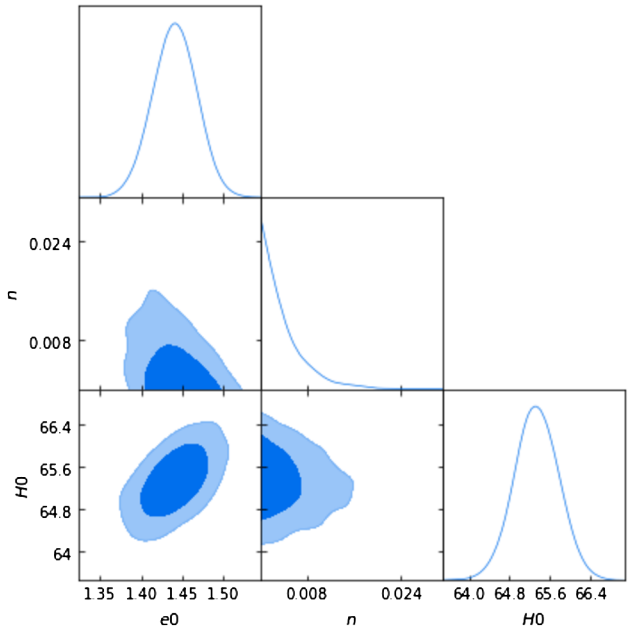


Fig. 4 The contour maps of $\bar{\zeta_0}$, n, H_0 at $1\sigma(68.3\%)$ and $2\sigma(95.4\%)$ confidence level using $SNe + OHD + BAO$. In Fig. $e0$ means $\bar{\zeta_0}$

Table 2 The transition scale factor and redshift, and present values of deceleration parameter and equation of state parameter using best-fit results of model parameters

Data	a_T	z_T	q_0	w_{eff}
SNe+OHD	0.504	0.984	−0.363	−0.575
SNe+OHD+BAO	0.660	0.515	−0.223	−0.482

erated phase to accelerated phase are listed in Table 2. We plot the deceleration parameter versus redshift in Fig. 5 for best-fit values of model parameters. Both the trajectories show that there is transition from decelerated phase to accelerated phase. In Fig. 5, red dotted line represents trajectory obtained by $SNe + OHD$ where as the blue solid line represents the trajectory from $SNe + OHD + BAO$. It is observed that the deceleration parameter of this viscous model starts from negative redshift, $z = -1$ and takes the transition from negative to positive. From (24) we can obtain the value of q_0 as $q(z = 0, \bar{\zeta_0}) = (1 - n - \bar{\zeta_0})/2$. Using the best-fit values of parameters from Table 1, the deceleration parameter today are listed in Table 2. This is comparatively higher from the observational results on the present value of q, which is around -0.64 (Tegmark et al. 2006; Hinshaw et al. 2009). The transition redshift, at which q enters the negative value region, corresponding to an accelerating epochs are also listed in Table 2. An analysis of the ΛCDM model with combined $SNe + CMB$ data gives the transition redshift as $z_T = 0.45 - 0.73$ (Alam et al. 2004).

Using (29) and (30), the trajectories of $\{r, s\}$ in $s - r$ plane for different values of n and $\bar{\zeta_0}$ are shown in Fig. 6. It is observed that in $0 \leq \bar{\zeta_0} \leq (3 - n)$, the trajectories start from quintessence region ($r < 1$, $s > 0$) in early time and tend to ΛCDM in late time. However, for the values in $\bar{\zeta_0} \geq (3 - n)$, the trajectories start from chaplygin gas region ($r > 1$, $s < 0$). It reveals that the (r, s) trajectory for best fitted parameters are lying in the region of quintessence. In all cases we find that the viscous model tends to ΛCDM model in late time. So the present model resembles the ΛCDM model in future. The present position of the universe dominated by the viscous matter corresponds to $\{r_0, s_0\} = \{0.43706, 0.2595\}$. Thus, the viscous model is distinguishably different from the ΛCDM model.

The $r - q$ relation for the model is shown in Fig. 7. The $SCDM$ model and steady state (SS) model correspond to

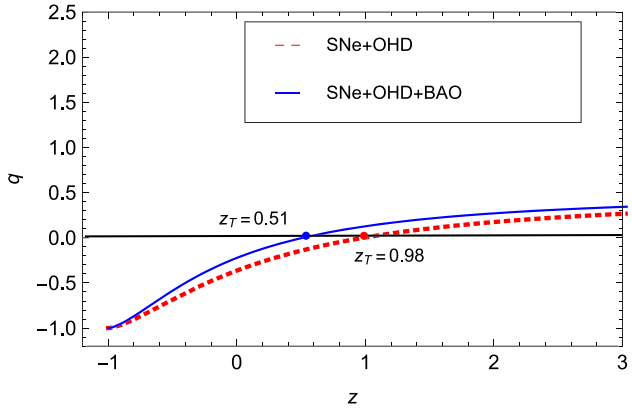


Fig. 5 Plot of the deceleration parameter q with redshift z for best-fit values of model parameters obtained from two sets of joint analysis of observational data

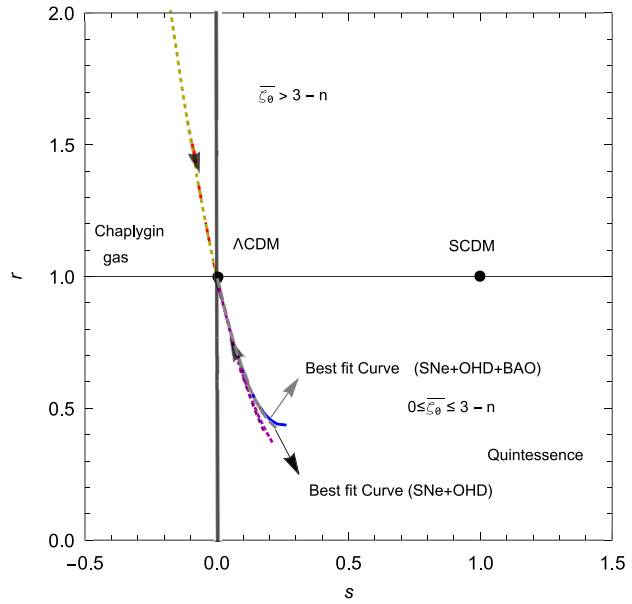


Fig. 6 The trajectory of $\{r, s\}$ in $s - r$ plane for the best-fit values and other different values of model parameters. The arrow shows the direction of the evolution of the trajectory

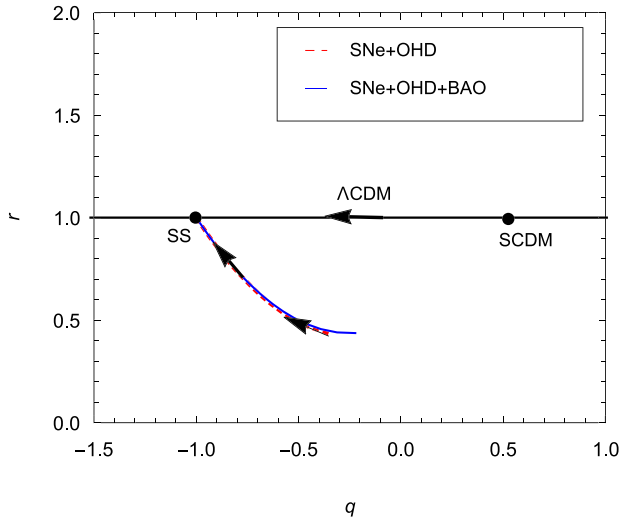


Fig. 7 The trajectory of $\{r, q\}$ in $q - r$ plane for the best-fit values of model parameters obtained from two sets of joint analysis of observational data. The arrow shows the direction of the evolution of the trajectory

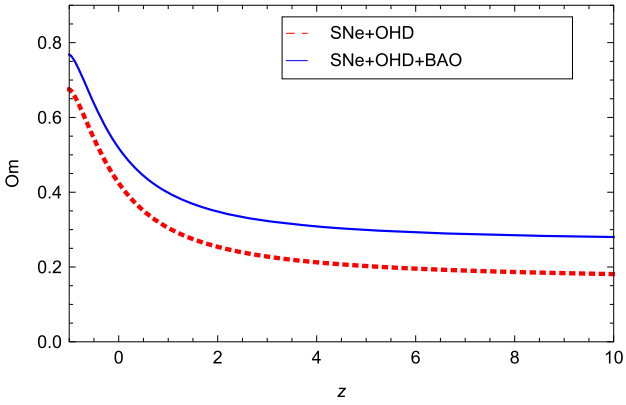


Fig. 8 The trajectory of $Om(z)$ for the best-fit values of parameters obtained from two sets of joint analysis of observational data

fixed points $\{r, q\} = \{1, 0.5\}$ and $\{r, q\} = \{1, -1\}$, respectively. The horizontal line at $r = 1$ corresponds to the time evolution of ΛCDM model. The viscous model approaches to the standard model like ΛCDM and quintessence model (Sahni et al. 2003) in late time.

Using Table 1 into (32), the trajectory of Om with redshift z is shown in Fig. 8. The trajectories reveal that the viscous model shows deviations from ΛCDM model which has horizontal line. Moreover, the negative slope of the trajectories indicate that our model has a quintessence behavior.

Let us calculate the effective equation of state (EoS) parameter to observe the accelerated expansion of the universe. It is to be noted that an accelerated expansion is possible only if the effective EoS parameter $\omega_{eff} < -1/3$, or $3\omega + 1 < 0$. The EoS parameter can be obtained using the

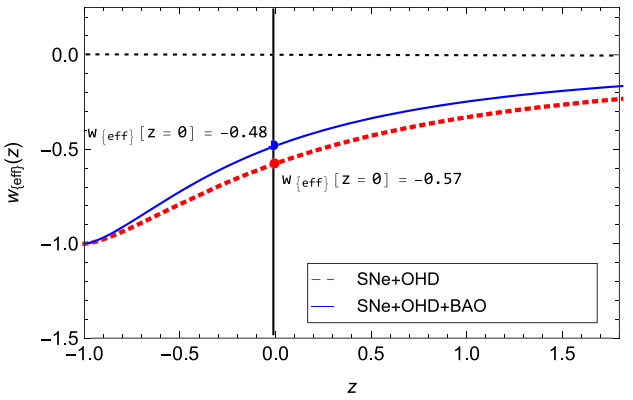


Fig. 9 Plot of effective equation of state parameter ω_{eff} with respect to the redshift z for best-fit values of model parameters obtained from two sets of joint analysis of observational data. The dot denotes the present value of ω_{eff} which is within the quintessence region

standard relation (Praseetha and Mathew 2014) as

$$\omega_{eff} = -1 - \frac{1}{3}\frac{2a}{h}\frac{dh}{da}, \tag{40}$$

where $h = H/H_0$ is the weighted Hubble parameter. Using (20) into (40), we get

$$\omega_{eff} = -1 + \frac{1}{3}\frac{(3 - n - \bar{\zeta_0})(1+z)^{\frac{(3-n)}{2}}}{h} \tag{41}$$

As $z \to -1$, $(a \to \infty)$, we get $\omega_{eff} \to -1$, which can also be observed from Fig. 9. It means that the model corresponds to de Sitter in future time. The EoS does not cross the phantom divide line $\omega \leq -1$ which shows that the viscous model is free from big-rip singularity.

The present value $(h = 1)$ of ω_{eff} is found to be

$$\omega_{eff}(z = 0) = -1 + \frac{(3 - n - \bar{\zeta_0})}{3}. \tag{42}$$

This shows that $\omega_{eff}(z = 0)$ makes transition from positive values to negative values for $\bar{\zeta_0} \geq (1 - n)$. It can also be understood by the present values of $q_0(z = 0)$, where q makes the transition for $\bar{\zeta_0} \geq (1 - n)$, which is discussed in Sect. 3. As stated above, $\omega_{eff} < -1/3$ is the condition for accelerating universe. From (42), it is obvious that $\omega_{eff}(z = 0) < -1/3$ only for $\bar{\zeta_0} \geq (1 - n)$, i.e., $(1 - n) \leq \bar{\zeta_0} \leq (3 - n)$. The present value of the EoS parameter by using the best-fitted values of model parameters obtained from both joint data, is listed in Table 2. These values are comparatively higher than that predicted by the joint analysis of $WMAP + BAO + H_0 + SNe$ data, which is around -0.93 (Komatsu et al. 2011). The evolution of effective EoS with redshift for the best estimated values of model parameters is shown in Fig. 9. The trajectories show that the viscous BD model favors the quintessence type of DE.

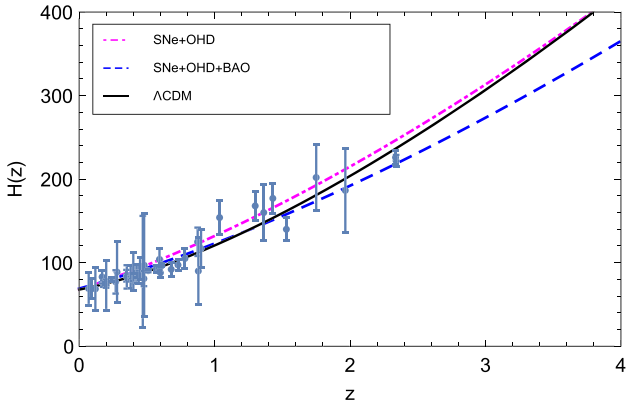


Fig. 10 Plot of Hubble evolution of the ΛCDM model and the best-fit values of model parameters obtained from two sets of joint analysis of observational data with error bar plots from Hubble data

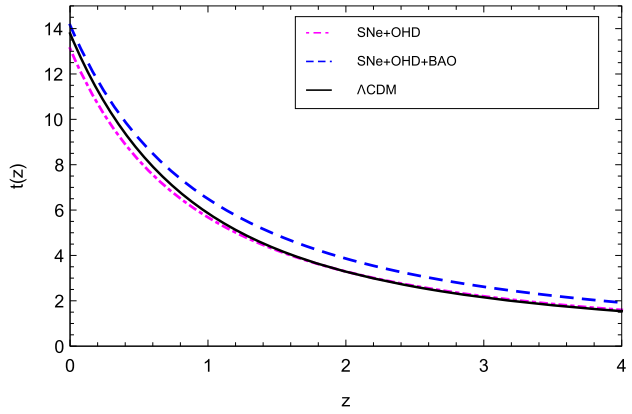


Fig. 11 The age of universe of viscous model for best-fit values of model parameters obtained from two sets of joint analysis of observational data and ΛCDM model in the units of Gyr with redshift

We compare our viscous model with the ΛCDM model with the error bar plots of Hubble dataset as shown in Fig. 10. We conclude that the fitting achieved from our statistical analysis is compatible with the Hubble observational data.

Let us calculate the age of the universe. The age of our universe at redshift z is given by $t(z) = T(z)/H_0$, where

$$T(z) = \int_z^\infty \frac{dz'}{(1+z')(H(z)/H_0)} \tag{43}$$

is the dimensionless age parameter. For the ΛCDM model, in which the density parameter, $\Omega_{m0} \approx 0.27$, the age parameter is (Feng and Li 2009)

$$T(z) = \int_z^\infty \frac{dz'}{(1+z')[\Omega_{m0}(1+z)^3 + (1-\Omega_{m0})]^{1/2}} \tag{44}$$

It is to be noted that the present age of the universe for a flat CDM model dominated by matter ($\Omega_{m0} = 1$, $t_0 = 2/(3H_0)$) gives $t_0 \approx 8$–10 Gyr, which does not satisfy the stellar age bound: $t_0 > 11$–12 Gyr, namely the age of the universe should be longer than any objects living in the universe. Therefore, this model suffers the age problem. However, for ΛCDM model, it easily satisfies the constraint $t_0 > 11$–12 Gyr. It seems that the problem is caused by the fact that matter is diluted too fast. Feng and Li (2009) claimed that the age problem would resolve once the viscosity is taken into account. In our viscous model, Fig. 11 shows that the trajectory for the Viscous model deviates very slightly from that of ΛCDM. It indicates that the viscosity could really alleviate the age problem. This is due to the change in the energy conservation equation by addition of viscosity which makes the matter dilution little bit slower, and then the age problem is alleviated.

7 Model selection

Based on statistical analysis we can determine which models are "better" taking into account how many parameters do the models need and how well do they fit the data. In this work we use two criteria, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). For a cosmological model with d degrees of freedom in which N number of data points have been used to fit the model, the AIC parameter is defined through the relation (Akaike 1974)

$$AIC = -2\ \ln \mathcal{L}_{\max} + 2d, \tag{45}$$

where $\mathcal{L}_{\max} = e^{-\chi^2_{tot}/2}$ is the maximum likelihood obtained for the cosmological model. The "preferred model" for this criterion is the one with the smaller value of AIC. To compare the model k with the model l, we calculate $\Delta AIC_{kl} = AIC_k - AIC_l$, which can be interpreted as "evidence in favor" of the model k compared to the model l. For $0 \leq \Delta AIC_{kl} < 2$ we have "strong evidence in favor" of model k, for $2 < \Delta AIC_{kl} < 4$, we have "average evidence in favor" of model k, for $4 < AIC_{kl} \leq 7$ there is "less evidence in favor" of the model k, and for $\Delta AIC_{kl} > 10$ there is basically "no evidence in favor" of model k (Liddle 2007).

On the other hand, the bayesian criterion is defined through the relation (Schwarz 1978)

$$BIC = -2\ \ln \mathcal{L}_{\max} + d\ \ln N, \tag{46}$$

where N is the number of data points. Similarly to ΔAIC_{kl}, $\Delta BIC_{ij} = BIC_i - BIC_j$ can be interpreted as "evidence against" the model i compared to the model j. For $0 \leq \Delta BIC_{ij} < 2$ there is "not enough evidence against" the model i, for $2 \leq \Delta BIC_{ij} < 6$ there is "evidence against" the model i and for $6 \leq \Delta BIC_{ij} < 10$ there is "strong evidence against" model i (Liddle 2007).

Using (45) and (46), it is obvious that a model giving a small AIC and a small BIC is favoured by the observations.

Table 3 Summary of the AIC and BIC for ΛCDM model and viscous BD model

Model	Data set	$\chi^2_{\min}$	AIC	BIC
ΛCDM				
	SNe+OHD	27.89	31.89	36.49
	SNe+OHD+BAO	30.45	34.45	39.21
Viscous BD				
	SNe+OHD	34.81	38.81	43.41
	SNe+OHD+BAO	72.67	76.67	81.43

It requires a 'reference model' for comparison, and ΛCDM could be the best reference model and we choose that for the purpose. Table 3 presents the results of AIC and BIC of ΛCDM model and viscous BD model.

In our analysis ΛCDM is the model with lower AIC and BIC parameters when we use data from the joint analysis of $SNe + OHD$ and $SNe + OHD + BAO$, respectively. From Table 3, considering ΛCDM as the referring model, we find $\Delta AIC = 6.92$ corresponds to $SNe + OHD$ and $\Delta AIC = 42.22$ corresponds to $SNe + OHD + BAO$. This indicates that the viscous BD model with joint analysis of SNe and OHD is in the region of "less evidence in favor" where as viscous model using joint analysis of SNe, OHD and BAO is in "no evidence in favour". Regarding BIC, we find $\Delta BIC = 6.92$ corresponds to $SNe + OHD$ and $\Delta BIC = 42.22$ corresponds to $SNe + OHD + BAO$. This shows that the model has "strong evidence against". We notice a tension between AIC and BIC results, while AIC indicates there is "less evidence in favor" with $SNe + OHD$, BIC indicates that there is "strong evidence against" for the same model. This is due to the fact that BIC strongly penalizes model. We can say that the viscous BD model obtained from joint analysis of SNe and OHD is marginally better than other results.

8 Conclusion

In this paper, we have discussed a bulk viscous matter-dominated Universe with constant bulk viscous coefficient, $\zeta = \zeta_0$ in Brans-Dicke theory. We have solved the evolution equations to obtain the scale factor and other cosmological parameters. We have identified three possible conditions for bulk viscous coefficient, $0 \leq \bar{\zeta_0} < (3-n)$, $\bar{\zeta_0} = (3-n)$ and $\bar{\zeta_0} > (3-n)$. Depending on these conditions, we have analyzed the different possible evolutions predicted by the viscous model. We have observed that the viscous model predicts a Universe with a Big-bang in the past and then transition from decelerating epoch to an accelerating epoch in late times for the values of $\bar{\zeta_0}$ in the range $0 < \bar{\zeta_0} \leq (3-n)$. For $\bar{\zeta_0} = (3-n)$, the model corresponds to the de Sitter and it always predicts accelerated expansion for $\bar{\zeta_0} > (3-n)$. We have calculated the value of the scale factor where the transition happens. We have found that the transition from decelerated phase to accelerated one occurs in future in $0 < \bar{\zeta} < (1-n)$. When $\bar{\zeta_0} = (1-n)$, the transition from decelerated to accelerated epoch takes place today. The transition between the deceleration-acceleration epoch takes place in past in $(1-n) < \bar{\zeta_0} \leq (3-n)$. In limiting case, when $\bar{\zeta_0}$ approaches to $(3-n)$, the transition takes place very close to Big-bang. The evolutions of the scale factor are plotted in Figs. 1 and 2 for different values of $\bar{\zeta_0}$ and n.

We have constrained the present viscous model using two sets of combined observational data coming from $SNe + OHD$ and $SNe + OHD + BAO$. The value of bulk viscous coefficient is positive which satisfies the second law of thermodynamics. We have shown the contour maps of parameters H_0, n and $\bar{\zeta_0}$ with $1\sigma(68.3\%)$ and $2\sigma(95.4\%)$ confidence level in Figs. 3 and 4.

We have studied the evolution of the deceleration parameter which is found to be time-dependent. We have calculated the present value of $q_0 = (1-n-\bar{\zeta_0})/2$. The present Universe enters into accelerating phase at an early stage, corresponds to $q_0 < 0$ for $\bar{\zeta_0} > (1-n)$. It enters into accelerating epoch in future time if $0 < \bar{\zeta_0} < (1-n)$, and the accelerating phase occurs at present time if $\bar{\zeta_0} = (1-n)$, i.e., $q_0 = 0$. The variations of q with z are shown in Fig. 5 for two set of joint analysis of model parameters. It is found that for the best estimates, the Universe enters into accelerating phase in recent past at a redshift $z_T = 0.984$ using joint observational data set of SNe and OHD where as it is $z_T = 0.515$ using $SNe + OHD + BAO$.

We have analyzed the statefinder and Om diagnostic parameters to discriminate our model from the other existing models especially from ΛCDM model. In Fig. 6, the trajectory of $\{r, s\}$ in $s - r$ plane for best fitted value of model parameters has been plotted. It reveals that the trajectory starts from quintessence region and tends to ΛCDM model in late times. The present value of $\{r, s\}$ in the $s - r$ plane corresponds to $\{r_0, s_0\} = \{0.43706, 0.2595\}$ using joint data of SNe, OHD and BAO which shows that the viscous model is distinguishably different from the ΛCDM model. However, as $a \to \infty$, $\{r, s\} \to \{1, 0\}$ corresponds to ΛCDM point. In Fig. 8, negative slope of $Om(z)$ trajectories also reveal that our model has quintessence behavior.

We have discussed the effective EoS parameter for the best estimates of the viscous model. It is observed that $\omega_{eff} \to -1$ as $z \to -1$, which shows that the viscous model behaves like de Sitter in future. The model has no big rip singularity as ω_{eff} does not cross the phantom divide line as shown in Fig. 9. The present value of ω_{eff} is around $\omega_{eff} \simeq -0.575$ through $SNe + OHD$ and $\omega_{eff} \simeq -0.482$ through $SNe + OHD + BAO$ for the best-fit values of parameters.

We have compared our model with the ΛCDM model with the error bar of Hubble dataset as shown in Fig. 10. It

is found that the fitting achieved our statistical analysis is compatible with the observational data used. We have also calculated the age of the universe and plotted $t(z)$ versus z in Fig. 11. It is found that the trajectories of viscous model deviate very slightly from that of ΛCDM model in $t_0 > 13 - 14\ Gyr$. Thus, the viscous model in BD theory alleviates the age problem.

In Sect. 7, we have performed the information criterion of *AIC* and *BIC* to discriminate our model with ΛCDM model. The analyses based on the *AIC* indicates that there is less support for the viscous model when compared to the ΛCDM model, while those based on the *BIC* indicates that there is strong evidence against it in favor of the ΛCDM model.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Ade, P.A.R., et al.: Astron. Astrophys. **594**, A13 (2016)
Akaike, H.: IEEE Trans. Autom. Control **19**, 716 (1974)
Alam, U., Sahni, V., Starobinsky, A.A.: J. Cosmol. Astropart. Phys. **0406**, 008 (2004)
Anderson, L., et al.: Mon. Not. R. Astron. Soc. **441**, 24 (2014)
Avelino, A., Nucamendi, U.: AIP Conf. Proc. **1026**, 300 (2008a)
Avelino, A., Nucamendi, U.: AIP Conf. Proc. **1083**, 1 (2008b)
Avelino, A., Nucamendi, U.: J. Cosmol. Astropart. Phys. **04**, 06 (2009)
Bamba, K., Odintsov, S.D.: Eur. Phys. J. C **76**, 18 (2016)
Banerjee, N., Pavon, D.: Phys. Lett. B **647**, 447 (2007)
Bennet, C.L., et al.: Astrophys. J. Suppl. Ser. **148**, 1 (2003)
Bertotti, B., Iess, I., Tortora, P.: Nature **425**, 374 (2003)
Betoule, M., et al.: Astron. Astrophys. **568**, A22 (2014)
Beutler, F., et al.: Mon. Not. R. Astron. Soc. **416**, 3017 (2011)
Blake, C., et al.: Mon. Not. R. Astron. Soc. **425**, 405 (2012)
Brans, C.H., Dicke, R.H.: Phys. Rev. **124**, 925 (1961)
Brevik, I., Grøn, Ø.: Relativistic Viscous Universe Models p. 99. Nova Sci. Publ., New York (2013). arXiv:1409.8561
Brevik, I., Hallanger, A.: Phys. Rev. D **69**, 024009 (2004)
Brevik, I., Odintsov, S.D.: Phys. Rev. D **65**, 067302 (2002)
Brevik, I., Gorbunova, O., Saez-Gomez, D.: Gen. Relativ. Gravit. **42**, 1513 (2010)
Brevik, I., Grøn, Ø., de Haro, J., Odintsov, S.D., Saridakis, E.N.: Int. J. Mod. Phys. D **26**, 1730024 (2017)
Caldwell, R.R.: Phys. Lett. B **545**, 23 (2002)
Caldwell, R.R., Kamionkowski, M., Weinberg, N.N.: Phys. Rev. Lett. **91**, 071301 (2003)
Cao, S.L., Teng, H.-Y., Wan, H.-Y., Yu, H.-R., Zhang, T.-J.: Eur. Phys. J. C **78**, 313 (2018)
Capozziello, S.: Class. Quantum Gravity **23**, 1205 (2006)
Capozziiello, S., Cardone, V.F., Elizalde, E., Nojiri, S., Odintsov, S.D.: Phys. Rev. D **73**, 043512 (2006)
Chiba, T., Okabe, T., Okabe, M.: Phys. Rev. D **62**, 023511 (2000)
Chimento, L.P., Richarte, M.G.: Eur. Phys. J. C **73**, 2352 (2013)
Clifton, T., Ferreira, P.G., Padilla, A., Skordis, C.: Phys. Rep. **513**, 1 (2012)
Copeland, E.J., Sami, M., Tsujikawa, S.: Int. J. Mod. Phys. D **15**, 1753 (2006)
Das, S., Banerjee, N.: Phys. Rev. D **8**, 043512 (2008)
Eckart, C.: Phys. Rev. **58**, 919 (1940)
Fabris, J.C., Goncalves, S.V.B., de Sa Ribeiro, R.: Gen. Relativ. Gravit. **38**, 495 (2006). astro-ph/0503362
Feng, C.J.: Phys. Lett. B **670**, 231 (2008)
Feng, C.J., Li, X.Z.: Phys. Lett. B **680**, 355 (2009)
Foreman-Mackey, D., Hogg, D., Lang, D., Goodman, J.: Publ. Astron. Soc. Pac. **125**, 306 (2012)
Giostri, R., et al.: J. Cosmol. Astropart. Phys. **1203**, 027 (2012)
Grøn, Ø.: Astrophys. Space Sci. **173**, 191 (1990)
Guo, Z.K., Piao, Y.S., Zhang, X.M., Zhang, Y.Z.: Phys. Lett. B **608**, 177 (2005)
Hinshaw, G., et al.: Astrophys. J. Suppl. Ser. **180**, 225 (2009)
Hinshaw, G., et al.: Astrophys. J. Suppl. Ser. **208**, 19 (2013)
Israel, W., Stewart, J.M.: Phys. Lett. A **58**, 213 (1976)
Kim, H.: Mon. Not. R. Astron. Soc. **364**, 813 (2005)
Komatsu, E., et al.: Astrophys. J. Suppl. Ser. **192**, 18 (2011)
Landau, L.D., Lifshitz, E.M.: Fluid Mechanics, vol. 6. Butterworth Heinermann Ltd., Oxford (1987)
Liddle, A.R.: Mon. Not. R. Astron. Soc. **377**, L74 (2007)
Maartens, R.: Class. Quantum Gravity **12**, 1455 (1995)
Murphy, G.L.: Phys. Rev. D **8**, 4231 (1973)
Nojiri, S., Odintsov, S.D.: Phys. Rev. D **72**, 023003 (2005)
Nojiri, S., Odintsov, S.D.: Phys. Lett. B **649**, 440 (2007)
Padmanabhan, N., et al.: Mon. Not. R. Astron. Soc. **427**, 2132 (2012)
Perlmutter, S., et al.: Nature **391**, 51 (1998)
Perlmutter, S., et al.: Astrophys. J. **517**, 565 (1999)
Pimentel, L.O.: Astrophys. Space Sci. **116**, 395 (1985)
Praseetha, P., Mathew, T.K.: Int. J. Mod. Phys. D **23**, 1450024 (2014)
Ren, J., Meng, X.H.: Phys. Lett. B **633**, 1 (2006)
Riess, A.G., et al.: Astron. J. **116**, 1009 (1998)
Sahni, V., Saini, T.D., Starobinsky, A.A., Alam, U.: JETP Lett. **77**, 201 (2003)
Sahni, V., et al.: Phys. Rev. D **78**, 103502 (2008)
Scherrer, R.J.: Phys. Rev. Lett. **93**, 011301 (2004)
Schwarz, G.: Ann. Stat. **6**, 461 (1978)
Setare, M.R.: Phys. Lett. B **654**, 1 (2007)
Sheykhi, A.: Phys. Rev. D **81**, 023525 (2010)
Singh, C.P.: Pramana J. Phys. **71**, 33 (2008)
Singh, C.P., Kumar, P.: Eur. Phys. J. C **74**, 3070 (2014)
Singh, C.P., Kumar, P.: Astrophys. Space Sci. **361**, 157 (2016)
Spergel, S., et al.: Astrophys. J. **148**, 175 (2003)
Spergel, S., et al.: Astrophys. J. **170**, 377 (2007)
Tawfik, A., Harko, T.: Phys. Rev. D **85**, 084032 (2012)
Tegmark, M., et al.: Phys. Rev. D **69**, 103501 (2004)
Tegmark, M., et al.: Phys. Rev. D **74**, 123507 (2006)
Velten, H., Wang, J., Meng, X.: Phys. Rev. D **88**, 123504 (2013)
Vikman, A.: Phys. Rev. D **71**, 023515 (2005)
Wang, J.X., Meng, X.H.: Mod. Phys. Lett. A **29**, 1450009 (2014)
Wei, H., Cai, R.G.: Phys. Lett. B **660**, 113 (2008)
Weinberg, S.: Rev. Mod. Phys. **61**, 1 (1989)
Will, C.: Living Rev. Relativ. **9**, 3 (2006)
Wu, F.Q., Chen, X.: Phys. Rev. D **82**, 083003 (2010)
Zhang, Y., Gui, Y.X.: Class. Quantum Gravity **23**, 6141 (2006)
Zhao, G.B.: Phys. Rev. D **72**, 123515 (2005)

Materials Research Express

ACCEPTED MANUSCRIPT

Reusing marble dust as reinforcement material for better mechanical performance: studies on compositing aluminum matrix

To cite this article before publication: Vipin Kumar Sharma *et al* 2020 *Mater. Res. Express* in press https://doi.org/10.1088/2053-1591/ab6702

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is **© 2020 IOP Publishing Ltd**.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.
As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by-nc-nd/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 14.139.251.100 on 08/01/2020 at 04:43

Reusing marble dust as reinforcement material for better mechanical performance: studies on compositing aluminum matrix

Vipin Kumar Sharma[1,2], Sumit Chaudhary[1], Vikas[1], Sonia[3], Ramesh Chandra Singh[1]

[1]Tribology Lab, Department of Mechanical Engineering, Delhi Technological University, Bawana Road, Delhi-110042, India

[2]Department of Mechanical and Automation Engineering, Maharaja Agrasen Institute of Technology, Rohini Sector-22, Delhi-110086, India

[3]Department of Production Engineering, Punjab Engineering College (Deemed to be University), Chandigarh- 160012, India

[*]**Corresponding Author e-mail**- vipin.dtu@gmail.com

Orcid Id: https://orcid.org/0000-0002-7356-4590 (Vipin Kumar Sharma)

Abstract

The present paper addresses the effects of the preheating of reinforcement particles on the performance of an aluminium metal composite. Waste marble dust particles were used as reinforcements in the Al 6063 aluminium matrix. The aluminium composites were fabricated by preheating the marbles particles at and 200 °C, 100 °C and at room temperature with the help of friction stir processing technique. Tensile strength, micro-hardness, wear and friction behavior of the fabricated materials were compared with aluminium marble dust composite fabricated with room temperature heated marble dust. It was observed that the aluminium marble dust composite processed with 200 °C preheated marble dust resulted in maximum tensile strength and micro-hardness as compared to the composite fabricated with room temperature heated marble dust. Improvement of 52% in tensile strength and 29% in micro-hardness as compared to the composite with room temperature preheated marble dust. The wear and friction properties for the Al-MD composite were also enhanced with the introduction of preheated marble dust in the aluminium matrix. There was an improvement of 25% in the specific wear rate and 23% in the coefficient of friction for the fabricated aluminium marble dust composite as compared to the composite with room temperature preheated marble dust.

Keywords: aluminium, composite, marble dust, wear, friction, materials research

Introduction

Metal matrix composites are used as a material for automobile. Aerospace and the other number of applications [Prasad and Shoba (2016)]. High hardness, strength, and wear resistance are some of the important properties of the composite materials for which these materials are successfully used in many industrial applications. Ceramic particles like Al2O3, TiC, SiC, MgO, Zircon, B4C and whiskers/fibers are generally used as reinforcement in the metal matrix to improve mechanical and tribological properties.

The selection of reinforcement parameters such as the amount of reinforcement, size, and shape of particles and mixing of reinforcement particles in the metal matrix greatly affects the performance of the composite materials [Mishra et al. (2015)]. The cost and processing of the reinforcement particles also play an important role in the selection of reinforcements and the ever-increasing

Accepted Manuscript

demand for minimum cost reinforcement gives rise the usage of industrial and agro wastes as a potential reinforcement in the metal matrix. With this the focus of the current researchers has been shifting towards the usage of agro and industrial wastes.

Flyash, rice husk ash, bamboo leaf ash, rock dust particles are some of the waste reinforcements which has been successfully used to fabricate the AMC [Prakash et al. (2016), Sharma et al. (2017)]. Out of these industrial and agro wastes, waste marble dust contributed to major environmental problems due to the generation of different stages of mining and related operations. The generation of marble waste starts from mining and lasts for the development of finished products. Leaving the waste particles in environment directly might cause environmental problems. Hence, many researchers are working on how to use this waste material in an effective way [Aliabdo et al. 2014)]. A brief review of the usage of waste marble dust in different applications is presented here.

Kumar(2018) used the marble dust powder as a possible reinforcement in the AlCuNi alloy. The marble dust was varied in the proportion of 3, 4, 9 and 12 % in the matrix material. It was reported by authors that, lower amounts of marble dust particles improved the tensile strength and hardness of the aluminium composite. With higher percentages (12%), the marble dust particles tend to agglomerate which in result decreased the mechanical properties. Ramesh et al. (2013) used the quarry dust particles to fabricate an AMC with the help of a stir casting method. Quarry dust in weight percentages of 5, 7.5 and 10 was used and the prepared composite was investigated to determine the effects of quarry dust particles on the mechanical properties. Increased strength, hardness and wear resistance was reported by the authors with the addition of quarry dust particles.

Rajak et al. (2019) proposed a copper-marble dust composite as a bearing material. Different Copper composites with marble dust (1.5,3.4.5 and 6 wt.%) particles were prepared using the stir casting in a vacuum. It was reported by authors that, marble dust enhanced the mechanical properties of the copper composite. The marble dust particles reduced the porosity in the composite and improved the bond strength between the matrix materials. Alaneme et al. (2018) manufactured Al-Mg-Si alloy from two-step stir casting method with groundnut shell ash and SiC to examine microstructural and mechanical properties, the author showed that ultimate tensile strength (UTS) and fracture toughness of the composite increased but hardness and strength decreased

Apart from the fabrication of aluminium composites, marble dust has been successfully used as an additive for the constructional bricks and concrete materials. With the addition of marble dust, mechanical, chemical and physical properties of the bricks get improved (Bilgin et al. (2012)) and marble dust can replace the clay in the bricks (Marras et al. (2010)). For concretes structures, marble particles enhanced the compressive strength and reduced the porosity of concrete (Demirel (2010)). Aliabdo et al. (2014) used the marble dust as a possible constituent of the cement and concrete construction. It is reported that the introduction of marble dust in the cement does not affect the initial and final setting time of the cement. The scanning electron microscope images also confirm the similarity in size for the marble dust and cement particles. The marble dust also helped in improving the compressive and tensile strength of the cement and concrete. For the best results, Aruntas et al. (2010) suggested using 10 wt% marble dust in cement preparation.

Accepted Manuscript

To utilize the benefits of two or more reinforcing agents in a single matrix, hybrid composites were produced by many researchers (Guru et al. (2007)). Marble dust along with flyash were mixed to prepare a polyester matrix composite. The experimental results concluded that the mixture of marble dust and flyash enhanced the strength and hardness of the composite. Chaithanyasai et al. (2014) investigated the effects of eggshell on the mechanical properties of Al 6061. Powder metallurgy route was adopted to fabricate the Al- eggshell composite. The authors reported good bonding between eggshell and Al powder and eggshells were properly distributed in the matrix of Al 6061. With the increase in wt% of eggshell increased the hardness of composite also increased.

Stir casting and powder metallurgy are the two famous methods which are used to fabricate the composite materials. The problem of agglomeration of reinforcement particles in the stir casting process and need of complex tooling for the powder metallurgy process are the two main factors which limit the usage of these methods to fabricate the finely structured surface composites. To develop fine structured surface composites, friction stir processing (FSP) is used. Numerous studies have been reported in the past by using the FSP method to fabricate the AMCs. The FSP process further breaks the reinforcement particles and improves the distribution of the reinforcing particles. Owing to this the typical casting effects get eliminated and strength and ductility get improved [Dinaharan et al. (2019)]. Another factor which improves the mechanical properties of the composite with FSP is the number of passes. With increase in number of passes to process a particular material. The microstructure becomes more compact which in result improves the tensile strength and hardness [Senthilkumar et al. (2019)].

Mehta et al. (2019) studied the wear behavior of reinforced Al-6061-T6 SiC made from FSP with boron carbide (B_4C). Two specimens were prepared with 1 pass for capping the groove and 3 were for stirring processing. In specimen-1, the direction of successive processing passes was the same. The sample was rotated 180° for successive processing passes, thus reversing the processing moves direction. From the experiment, the author observed that the specimen-1 wear was 72.15% less than the base metal, while specimen-2 showed 74.82% less wear than base metal and 9.58% less wear than specimen-1. Patel et al. (2017) studied the effect on tensile, compressive, yield strength, hardness, and toughness of reinforced AL6061 with Fly ash and E-glass fiber by stir casting method. The author saw that tensile strength, yield strength, and compressive strength, percent elongation, and hardness rises through the addition of F.A with continuous E-glass fiber but only the 6 percent fly ash sample decreases in hardness. Only by adding the fly ash reduces toughness. Prabhu et al. (2019) reported some improvement of the friction stir processed AA6082/$CaCO_3$ composite microstructure (10-12 μm) as a result of dynamic recrystallization. The reduction in grain size about 10-12 μm from 141 μm grain size, microhardness increased from 64 Hv to 87 Hv and the wear rate was reduced by 1/3 of AA6082 and 2/3 of FSP AA6082 was reported by the authors. Bannaravuri (2018) compared the effect on density, tensile, hardness, impact strength, tribology, and microstructure by fabrication of Al-4.5%Cu matrix alloy bamboo leaf ash by stir casting technique. The author discovered from this experiment that the bamboo leaf ash spread evenly throughout the matrix. Density reduces as the percentage wt of bamboo leaf ash rises, the strength of tensile and hardness increases, the rate of wear decreases as the percentage wt of bamboo leaf ash rises.

In most of the literature, waste marble dust has been used in the constructional materials. It is analyzed that marble dust particles in the aluminium matrix have proved to improve mechanical

Accepted Manuscript

properties, however, this area needs more explanation. The preheating of the marble dust particles also affects the performance of aluminium composites. Based on these observations, aluminium composite with marble dust particles as reinforcements has been fabricated using the friction stir processing technique. The effect of preheating of the marble dust particles on the characterization and mechanical properties were evaluated.

2. **Materials and Method**

2.1 Material selection

Al 6063 was used as the matrix material. The material was procured from Aluminium Trading Co., Delhi, India. The magnesium and silicon are the major alloying elements of it. The elemental composition of the Al 6063 as obtained by chemical spectroscopy is shown in Table 1. Various mechanical tests were carried out on Al 6063 to find out the mechanical properties. Table 2 presents the mechanical properties of Al 6063 material.

Table 1: Elemental composition of Al 6063 alloy

Elements	Si	Fe	Mg	Mn	Ti	Zn	Ni	Cr	Al
wt %	0.64	0.2732	0.6941	0.1074	0.0736	0.0693	0.039	0.0345	98.01

Table 2: Mechanical Properties of Al6063

Property	**Value**
Young's Modulus	68.3 GPa
Tensile Strength	165 MPa
Elongation	18-33%
Melting Temperature (Tm)	615^{o}C
Thermal Conductivity(k)	201-218 W/m*K

Marble stones were used as the reinforcing agent. Marbles stones are the natural stones having calcareous ($CaCO_3$) and dolomite CaMg (CO_3) major elements [El-sherbiny (2015)]. There were two types of waste obtained while cutting and polishing of marble. First waste was during the cutting and polishing of the marble dust was called marble waste slurry and the second waste was the broken pieces of marble. The total quantity of waste obtained while processing the blocks and slabs of marbles varies with the machining process was between 30% and 50% of the volume of all processed blocks. These wastes affect both nature and the economy and have harmful effects. The solid waste of marbles was stored in the huge factory warehouses. The marble slurry was refined as much as possible from the water through the treatment unit and the marble waste sludge was obtained. This amount of waste was too large for marble companies to be able to store. For this reason, with controlled or uncontrolled access, these marble wastes were dumped into nature. Because of this, the waste causes considerable pollution of the environment. This concern about the disposal of marble dust is the main basis of this research. The chemical composition of the

Accepted Manuscript

marble dust used to fabricate the composite material is shown in Table 3. From Table 3 it is revealed that the major portion of the marble dust is CaO, which was produced by the $CaCO_3$ upon heating during the testing. The conversion of $CaCO_3$ to CaO usually took place at higher temperatures. The presence of CaO in the marble dust shall be from the source material from where it was procured and therefore it is being detected during the testing of marble dust composition.

Table 3 Marble dust composition

Components	**Wt%**
CaO	42.45
MgO	1.52
SiO_2	26.35
Al_2O_3	0.520
Fe_2O_3	9.40

2.2 Experimental Setup

FSP machine was used for the processing of Al6063. Table 4 shows the full specifications of the FSP machine which was used for fabrication of composite. This machine consists of a 3-phase powerful motor that drives the FSP tool and provides an efficient torque to it. It has a hydraulic system which makes it efficient to use and work with accuracy. The hydraulic system was used to clamp the materials, up and down motion of the tool and the transverse motion of the base. The discharge of the hydraulic was controlled by the actuators using a nob. This affects the motion of the tool and the base plate.

Table 4: Friction stir processing setup specifications

Parameter	**Property**
Machine Size (L×B×H)	1300×1650×2000 mm
Welding Geometry	Straight
Machine Travel (x,y,z)	600,200,300 mm
Axis Thrust Force	250-2500 Kgf (x axis), 400-4000 Kgf (y-axis)
Spindle Speed	1440 rpm

2.2.1 FSP Tool

The FSP tool shape and material greatly affects the performance of the processing. FSP tool specifications and material properties were very critical to select and designing the tool-tip. The tool material basically depended on the application of tools and type of work-piece. The tool consists of a prob/tip, shoulder to generate frictional heat and apply the required downward force to consolidate and maintain the soften metal underneath of shoulder surface. Generally, the outer surface of shoulder consists of conical or cylindrical in shape. The diameter of the shoulder and tip were in a ratio of 3:1. The most common and straight forward design is the flat shoulder end

Accepted Manuscript

surface. This shoulder does not entrap the material under the surface and the unwanted material flashed out.
The tool pin was mainly responsible for distributing the contacting surface of the work-piece, cut the metal in face of the tool and move the plasticized metal behind the tool. The depth of processed zone and maximum traverse speed were controlled by the pin geometry. Various forms and characteristics, including threads, apartments or flutes, can be regarded on the exterior surface of the pin. Thread-less pins were better options for processing high strength or highly abrasive alloys as these characteristics can be readily worn away.

Table 5: Dimensions of workpiece and tool

Dimensions of Tool	
Tool Length (mm)	100 mm
Shoulder Diameter (mm)	19.95 mm
Pin Width (mm)	5 mm
Pin Diameter (mm)	7 mm
Pin type	Triangle
Dimensions of Plate	
Plate length (mm)	200
Plate width (mm)	75
Plate thickness (mm)	10
Dimensions of Groove	
Groove width (mm)	2
Groove depth (mm)	2
Groove length (mm)	200

3. Experimental Procedure

3.1 Friction Stir Processing

Friction Stir Processing was carried out with the help of the friction stir welding machine. The groove (2×2) was created on the Al6063 plate by the milling machine and the composite was fabricated, by using the FSP machine. The groove was properly cleaned by using acetone and the pre-heated marble dust particles were filled and closed by the closing method. The marble dust was preheated at a temperature 100ºC and 200 ºC temperatures. The preheating temperature of the marble dust was selected to make it sure to eliminate any amount of moisture content from the marble dust. In this work, the effects of preheating marble dust particles have been evaluated. An induction furnace was used to heat the marble particles [Rajak et al. (2019), Rajak and Aherwar (2016)]. The required amount of the marble dust was kept in an aluminium foil and kept in an induction furnace for 3 hours. The temperature of the induction furnance was closely observed during the preheating of the marble dust particles.

In the groove closing method, the groove was filled by marble dust and then it was closed by FSP probe-less tool to prevent escaping of the micro-particles. The principle of frictional heat was used for the closing of the groove. The frictional heat was generated by the phenomenon of the rubbing action between the tool probe-less tool and the workpiece. After the closing of the material, the

Accepted Manuscript

plate was subjected with the help of FSP tool with the probe. On the basis of literature review, FSP parameters were selected to carry out the experimental work. Table 6 shows the process parameters.

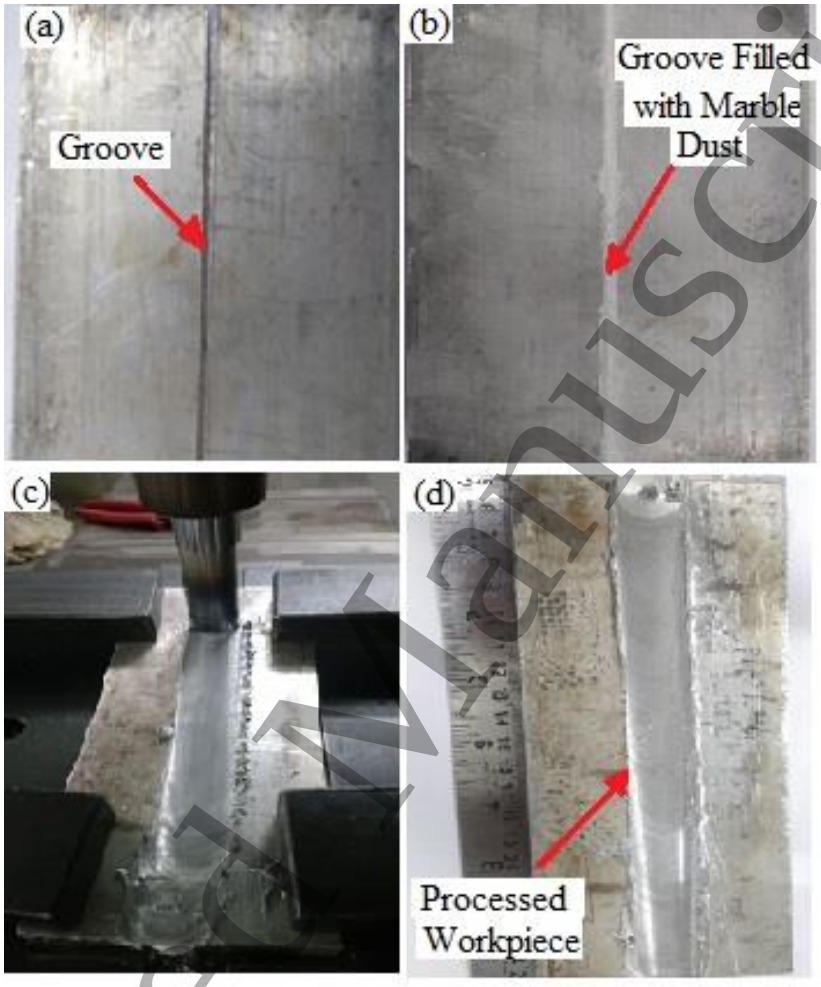


Figure 1: Step by step method to produce the composite material.

Table 6: Processing parameters used for the fabrication of composite material

Material	Feed rate (mm/min)	Tool angle (Degree)	Rotational speed (RPM)	Preheat Temp. (^{0}C)
Sample1 (Al 6063)	-	-	-	-
Sample 2 (Al-MD)	25	0^0	900	Room temp
Sample 3 (Al-MD)	25	0^0	900	100
Sample 4 (Al-MD)	25	0^0	900	200

3.2 Characterization and Mechanical Testing

3.2.1 Microstructure

The distribution of the reinforcement particles in the matrix material greatly affects the performance of the composite. An optical microscope was used to obtain the micrographs for the

Accepted Manuscript

fabricated composite materials. To obtain the image, the surface of the materials were polished using 100, 120, 220, 320, 400, 600, 1200, 2000 grits ambry paper. After that, wet polishing was done by contacting the samples with fine-grit emery paper against the rotating disc. Lastly, polishing was done on a proprietary cloth using alumina powder. Ethanol was used to wash the surface to remove contaminants. The samples were etched using Keller's reagent solution to unveil the grain structure. The microscope images were obtained at a magnification of 100X.

3.2.2 **X-ray diffraction (XRD)**

The XRD works on the principle of constructive interference of a monochromatic X-ray. The X-rays were focused and aimed at the prepared sample. The interaction of the incident ray with the sample results in constructive interference (and a diffracted ray) when conditions fulfill the law of Bragg. This law refers to the wavelength of electromagnetic radiation to the diffraction angle and the lattice spacing in a crystalline sample. The XRD was done for characterization of the composite material and the base metal to compare the mixing of reinforcement. After cutting the samples polishing was done with the help of fine emery paper to eliminate the roughness and to make even surface.

3.2.2 Tensile Strength

The tensile strength of the prepared composites was evaluated using a Tinius Olsen made ultimate tensile testing machine. The specimens for the tensile testing were obtained from the friction stir processed zone using the wire electric discharge machining as per the ASTM-E8. Figure 4 (a) presents the dimensional configuration of the tensile specimen and Figure (b-d) shows the actual specimens used to test the tensile strength. Each experiment was performed at a strain rate of 1mm/min and at room temperature. Three tests were formed for a single composite and average of three was considered to draw the results.

Accepted Manuscript

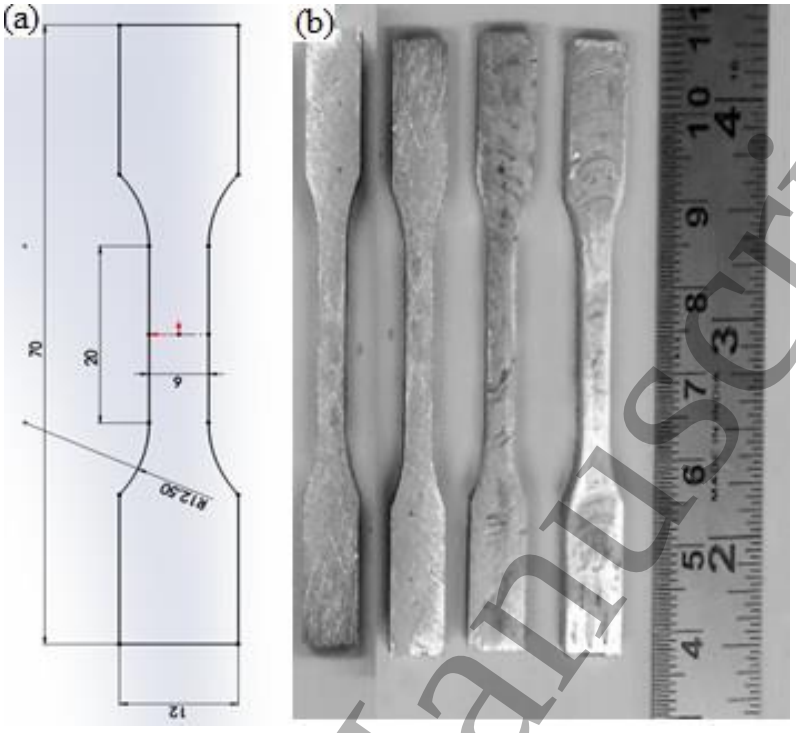


Figure 2: (a) Tensile Specimen design (b) Tensile specimen after wire cutting.

3.2.3 Micro Hardness

The hardness measurement was done as per ASTM E-384 standards. The Microhardness test was carried on Fischer made microhardness tester machine. Prior to the experiment, samples were polished from 100 to 1200 grit fine emery paper to remove the oxide layer and other scales from the surface. The Vickers microhardness values were taken using the load of 300mN with a dwell time of the 20s on the processed regions. For each specimen, ten observations were taken and the average of ten was used for analysis.

3.2.4 Tribological Testing

The wear and frictional behavior of the prepared specimens were evaluated using a pin on disc setup as per ASTM-G99 Standards. In pin on disc setup, the fabricated material was used as a pin which is slide against a disc. EN-31 steel (Vicker hardness number-746 HV) was used as the disc material.

For the testing, a pin on 8mm diameter and 4 mm height was cut from the friction stirred zone of the composite material. The pin was held in a mounting and later attached with a bell crack lever mechanism of the pin on disc setup. Dead weights were applied on the other side of the lever. Pin on disc setup had a friction sensor installed on it, which was used to measure the friction force between the pin and disc specimen, later the friction force was converted to the coefficient of friction.

The wear from the pin specimens was measured with the help of a precise weighing balance having an accuracy of 0.0001g. The specific wear rate was calculated from the wear obtained from the

Accepted Manuscript

pin. Table 7 presents the experimental parameters for the evaluation of wear and friction behavior. Figure 3 & 4 presents the test specimens before and after the wear test.

Figure 3: Wear test preparation

Table 7: Wear Parameters used for wear and friction testing

Parameters	Value
Disc temperature (ºC)	50
Disc Speed(RPM)	300
Sliding Distance(m)	500
Load(N)	40
Sliding distance (m)	500

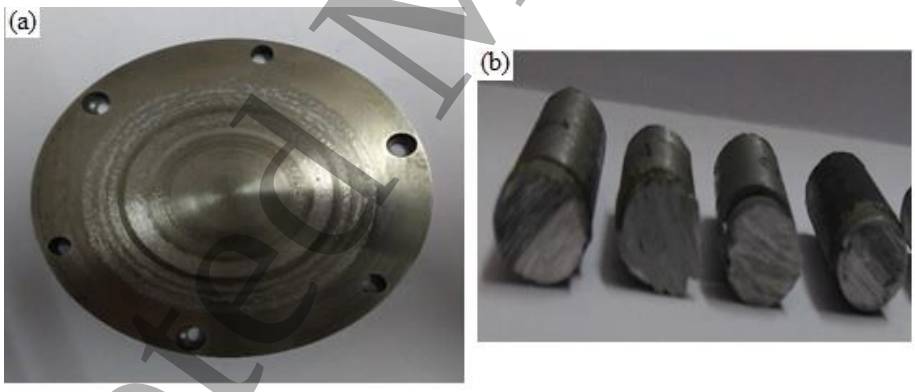


Figure 4: The worn pin and disc samples

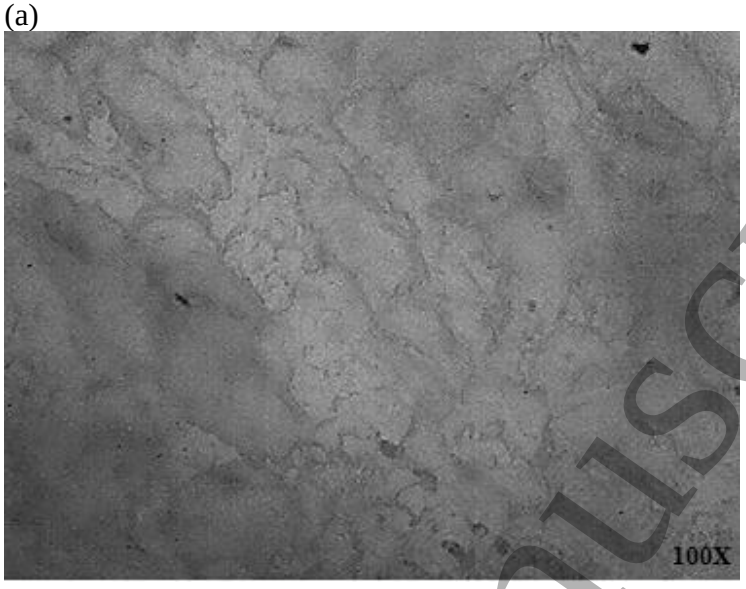
4 Results

4.1 Microstructure

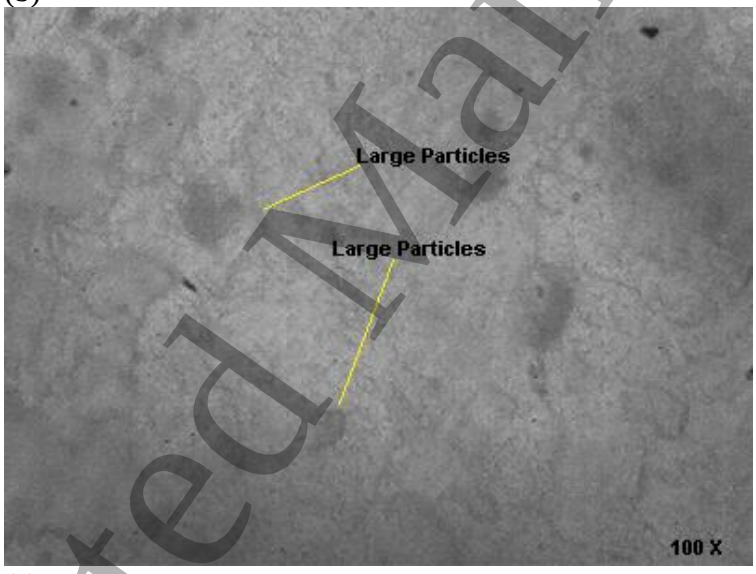
The optical micrographs obtained from the microscope are presented in Figure 5 (a-d). The presence of marble dust particles was detected for the prepared composites. The microstructure of the friction stir processed composites presents smaller grains as compared to the unprocessed Al 6063 (Figure 5 (a)). The grains were further found to get finer as the preheating temperature gets increased. Figure 5 (b), (c) and (d), show the microscopic images for composite prepared with marble dust heated to room temperature, 100℃ and 200℃ temperature respectively.

Accepted Manuscript

(a)

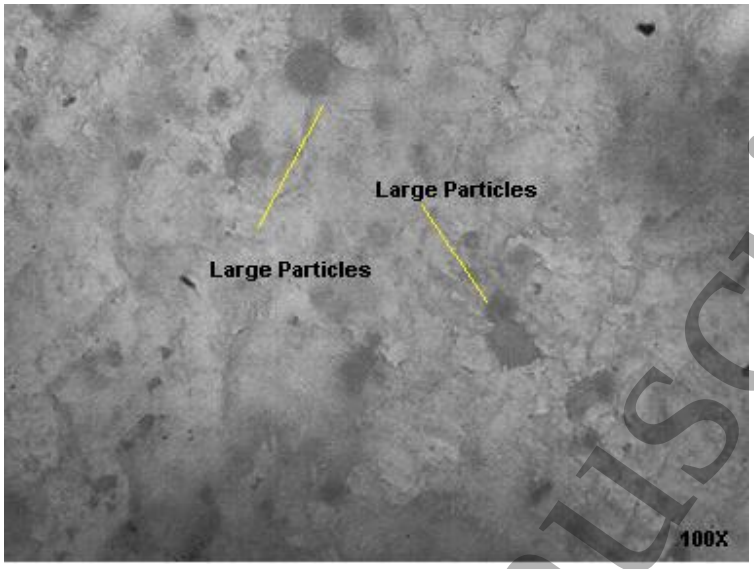


(b)



(c)

Accepted Manuscript



(d)

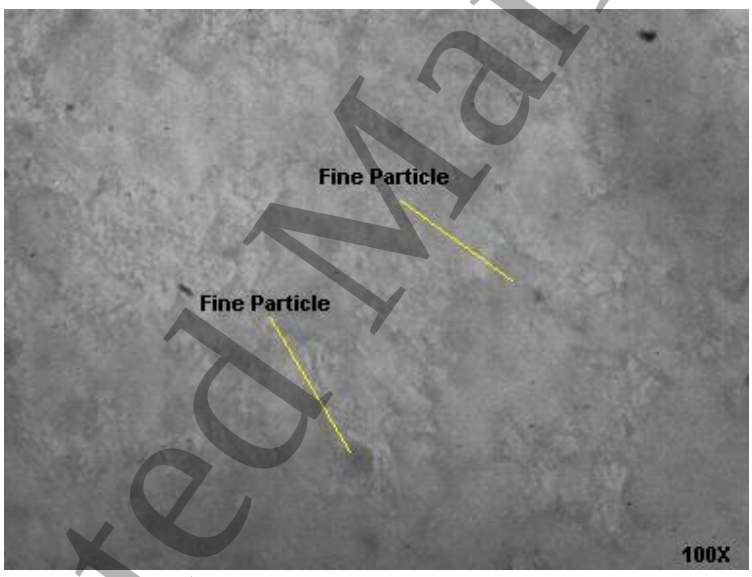


Figure 5: Optical micrographs for (a) unprocessed material Al 6063 (b) composite prepared with marble dust heat to room temperature (c) composite prepared with marble dust heat to 100 ℃ temperature (d) composite prepared with marble dust heat to 200 ℃ temperature

4.2 X-Ray Diffraction

The XRD was done for the detection of the intermetallic compound in the samples. The rays are subjected at a range of 2θ angle from 10º-80º throughout the analysis. The peaks are detected in this analysis. The results reveal the particles present in the sample of the Al6063 and in the processed samples. XRD results show that the major element in Al6063 was Al and Al-Fe-Si. The XRD of unprocessed Al6063 is shown in Figure 6.

Accepted Manuscript

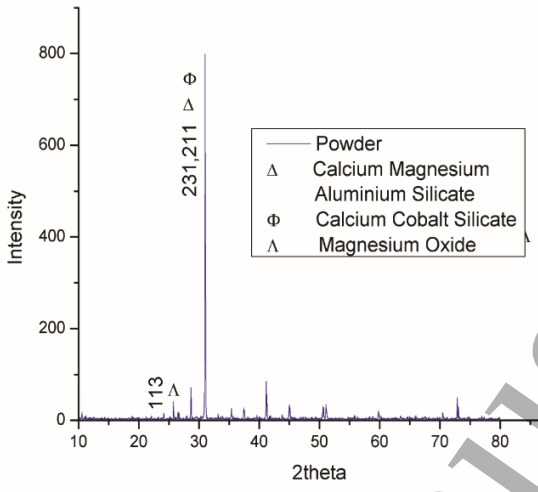
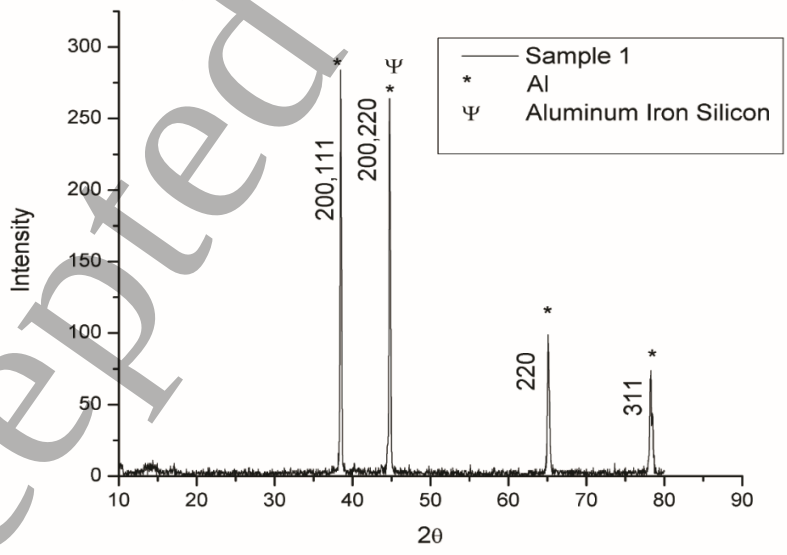


Figure 6: XRD of Marble dust

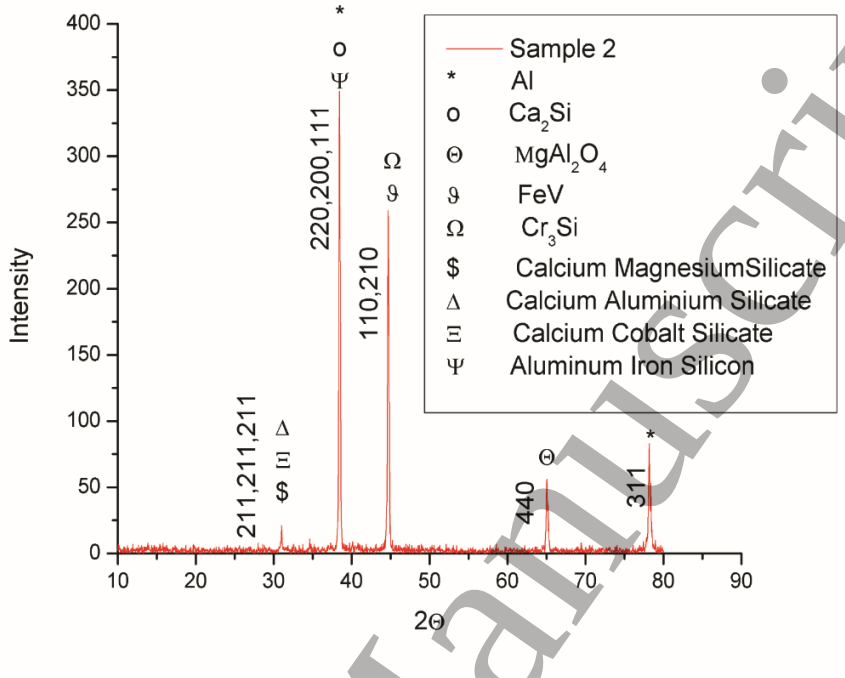
Figure 7 (a-d) shows the elements like Ca, Si, O, Al, and Mg which proves the presence of marble dust in the composite. The bonding between the M.D and the Al was shown in the microstructure. The XRD results also show that aluminum reacts with the atmosphere to form its oxide. Ca-Mg-Si was the main element formed in marble dust from the reaction. Ca-Mg-Si was seen at a 31^{o} 2θ angle.

(a)

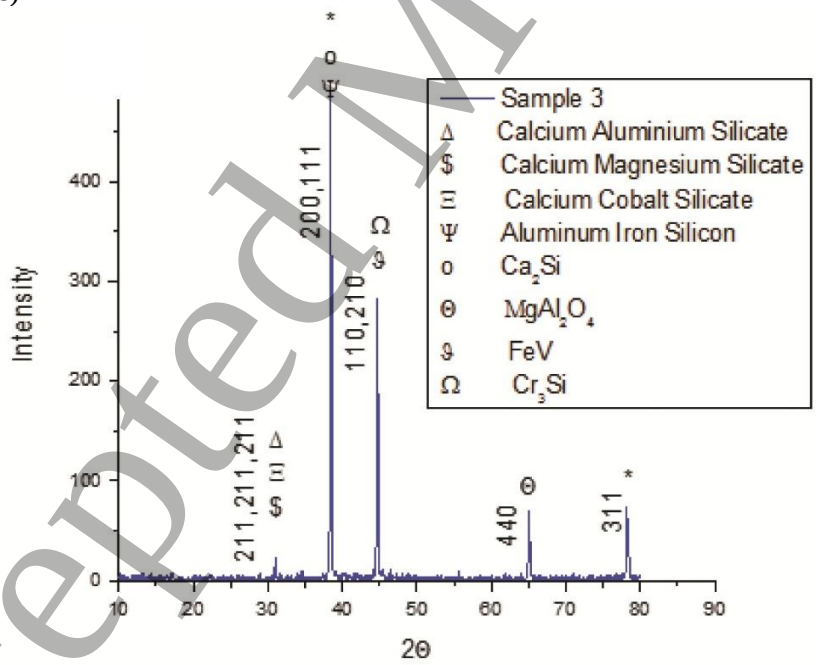


(b)

Accepted Manuscript



(c)



(d)

Accepted Manuscript

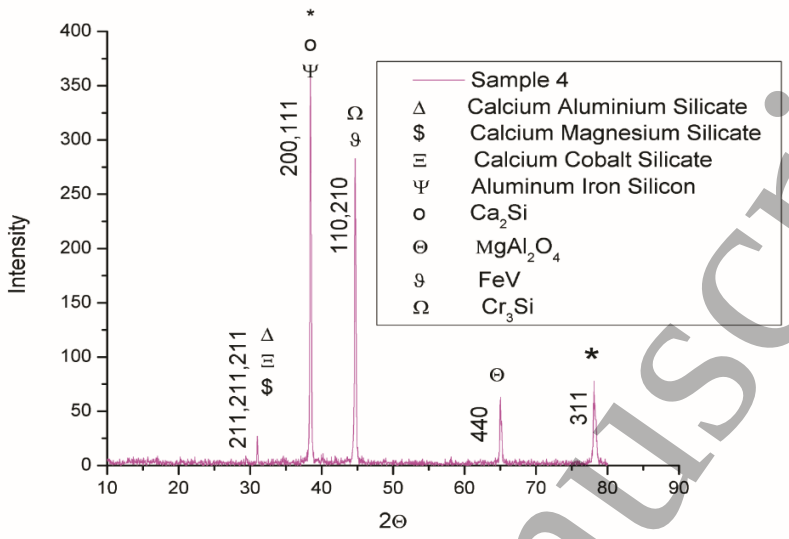


Figure 7: XRD results for (a) Al 6063 (b) composite prepared with marble dust heat to room temperature (c) composite prepared with marble dust heat to 100 ℃ temperature (d) composite prepared with marble dust heat to 200 ℃ temperature

4.3 Micro-hardness

The Vickers hardness of the specimen was calculated by taking 10 readings on the samples. The microhardness variation observed for the processed samples along the width is presented in Figure 8 (a). The AA 6063 aluminium had a Vicker hardness of 67 HV. The sample prepared with the marble dust heated to room temperature exhibits decrease in the hardness value with maximum hardness of 65 HV achieved at center of the processed zone. The low bonding between the Al and marble dust reduce the surface hardness of the processed sample. With increase in the preheating temperature to 100°C the maximum hardness value reached to 69 HV. The presence of CaO, SiO_2, MgO, and Al_2O_3 improved the hardness of the fabricated materials [Lakshmi et al. (2017), Prasat et al. (2011)]. The increased preheating temperature of the marble dust particles removed any leftover amounts of the moisture and friction stirring resulted in finer grain structure. With further increase in the preheating temperature, a compact finer grain structure at a higher temperature helped in further improving the micro-hardness. The better bonding strength of Aluminium with marble dust at higher preheating temperatures also enhanced the microhardness for the prepared composite [Rajak et al. (2019)].

The average micro-hardness values for the complete friction stirred zone are presented in Figure 8 (b). The average micro-hardness in the case of the composite prepared at 200°C pre-heat temperature was 77.396 HV, this increment was about 15% from the base material.

Accepted Manuscript

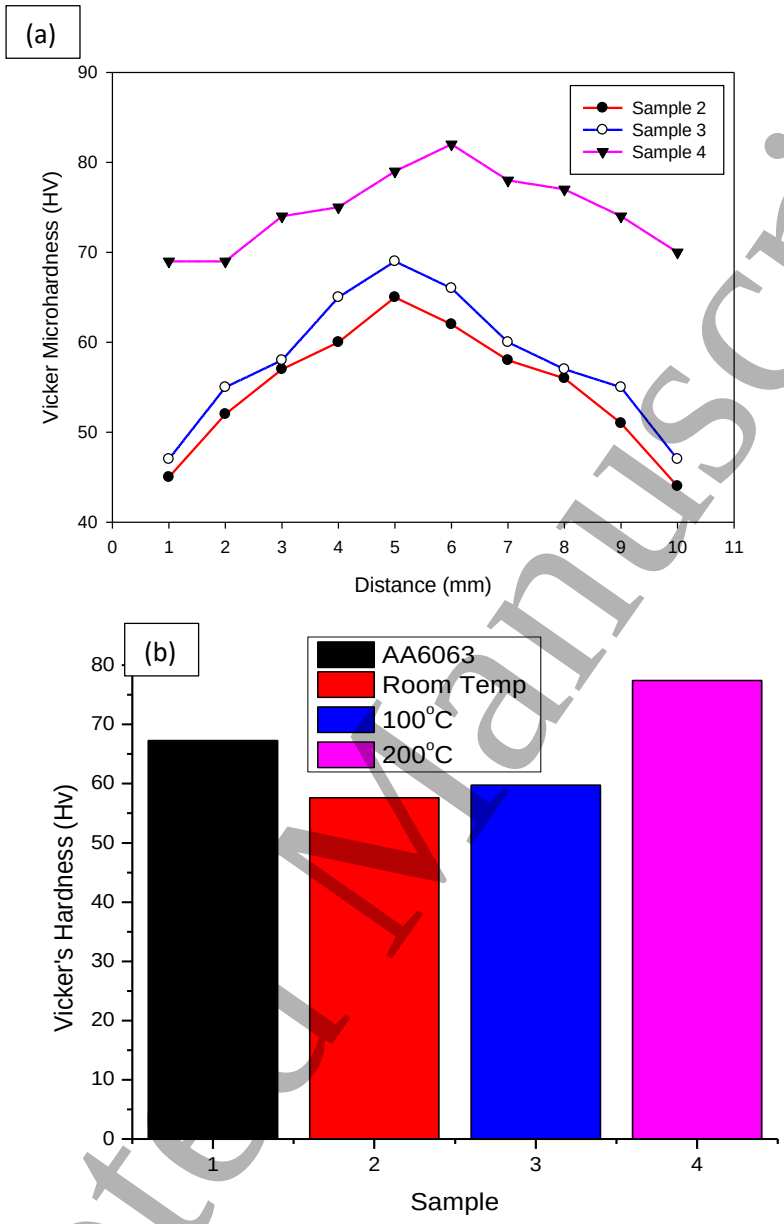


Figure 8: (a) Vicker's hardness along the processed zone (b) Average Vicker hardness for the prepared samples

4.4 Tensile Strength

The strength behavior of the processed samples is presented in Figure 9. The low bonding strength between the Al matrix and the marble dust particles preheated to room temperature and 100 ℃ resulted in some decreases in the tensile strength and ductility. The sample with preheated marble dust particle to 200℃ improved the strength of Al alloy. The tensile strength of Al 6063 alloy was 157 MPa and for sample 4 (pre-heated at 200℃) was 202 MPa. There was an increment tensile

Accepted Manuscript

strength of about 29 %. The tensile strength gets increased with increase in pre-heat temperature of the marble dust particles before FSP. The preheated (200°C) marble dust particle leads to an increase in tensile strength due to the excellent wettability and excellent bonding with the Al matrix. The refining of grains during the FSP and uniform distribution of marble particles in the Al 6063 matrix also helped in improving the tensile strength for the 200°C preheated marble dust mixed composite. The loads applied were transmitted in the composite to the marble dust particles, which increases the load-bearing capacity [Atuanya et al. (2012), Apasi et al. (2013), Hassan and Aigbodion (2015)]. Elongation was also increased from 6.5mm for Al6063 to 7.2mm for processed sample at 200°C. Table 8 shows the tensile strength and elongation values of all samples.

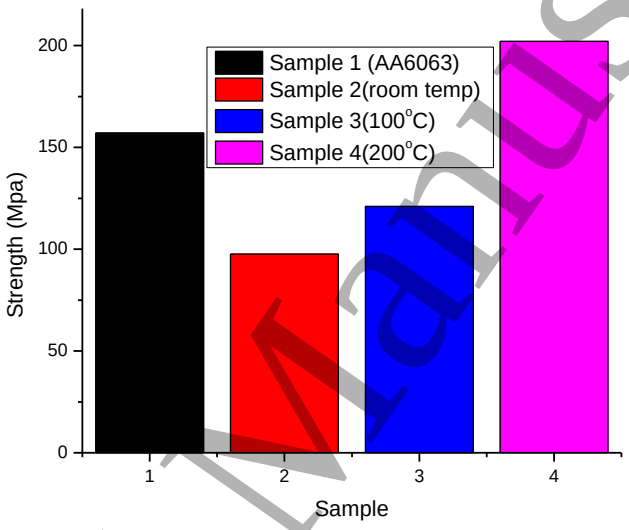


Figure 9: Variation of tensile strength for the fabricated composites.

Table 8: Strength and elongation of the specimen

Specimen	Strength (MPa)	Elongation(mm)
Sample-1	157	6.5
Sample-2	97.6	2.7
Sample-3	121	1.8
Sample-4	202	7.2

4.5 Tribology Testing

The tribological investigation was done on Al6063 and Al6063/marble dust composite using a pin on disc method [Sharma et al. (2018)]. The wear and friction properties of the material directly depends on the hardness and fracture toughness of the materials and tensile strength have limited effects on the wear and friction of the material [Alajmi and Shalwan (2015)]. This experimental study is limited to the evaluation of microhardness and tensile strength of the prepared material. For better correlation between the mechanical properties and wear behavior, the fracture toughness is need to be evaluated.

Accepted Manuscript

Figure 10 presents the variation of the specific wear rate for the processed samples. It is observed that sample 4 had the least amount of specific wear rate. The presence of finer grains in the aluminium matrix processed at 200℃ played an important role in decreasing the wear rate. The hardness of the samples and also the formation of the protective layer of Al_2O_3 in the presence of air helped in reducing the wear from the pin samples [Singh et al. (2019), Sharma et al. (2019)]. Sample 2 and sample 3 also reduces the wear rate.

The coefficient between the composite pins and EN-31 disc is shown in Figure 11, the presence of marble dust particles with compact structure at a higher temperature (sample 4) reduces the coefficient of friction between the tribopairs. This trend of variation of wear rate and coefficient of friction is similar as obtained by Rajak et al. (2019). Rajak et al. (2019) prepared the aluminium marble dust composites and investigated the wear and friction variations as one of the parameters.

Also, the trends obtained for variation of tensile strength and wear rate does not correlate. This presents the limited dependence of wear rate on the tensile strength of the composite materials. A similar correlation was reported by Alajmi and Shalwan (2015).

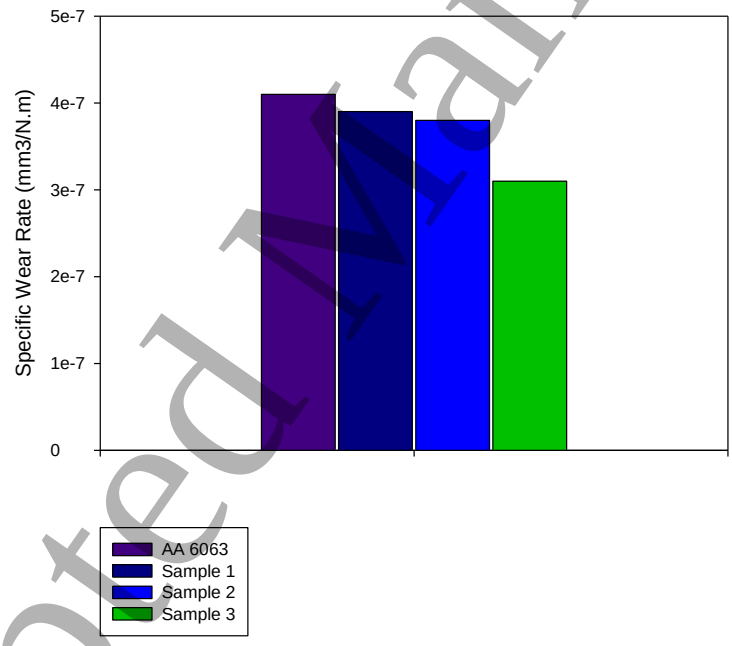


Figure 10: Specific wear rate variation for the test specimens

Accepted Manuscript

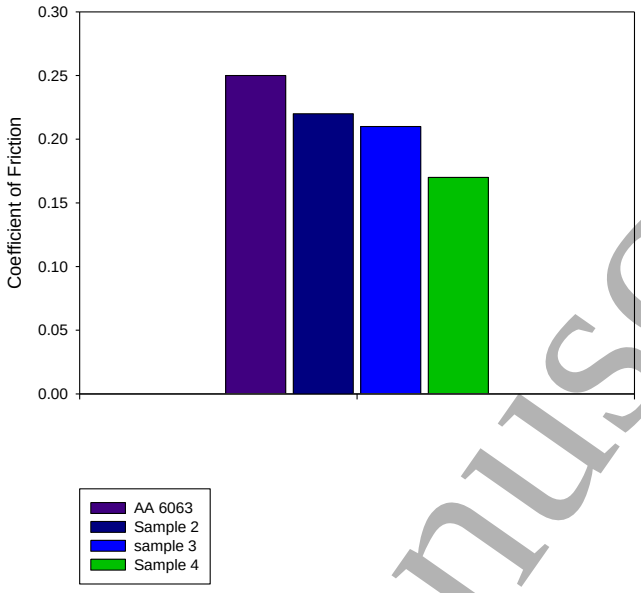
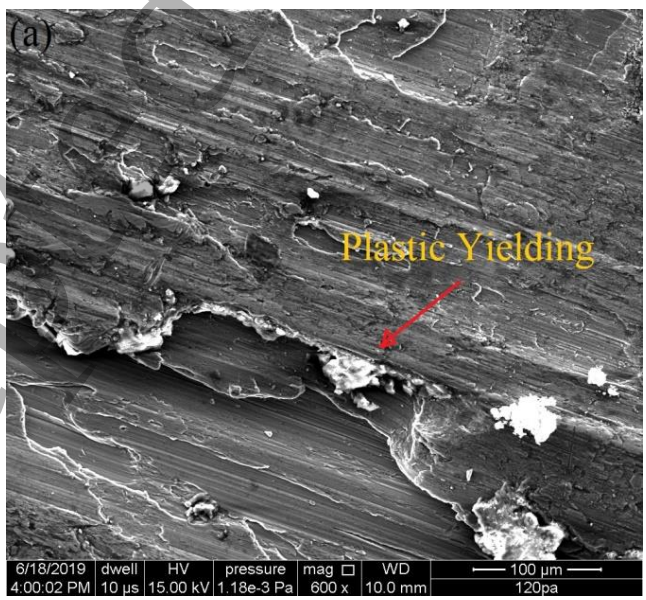
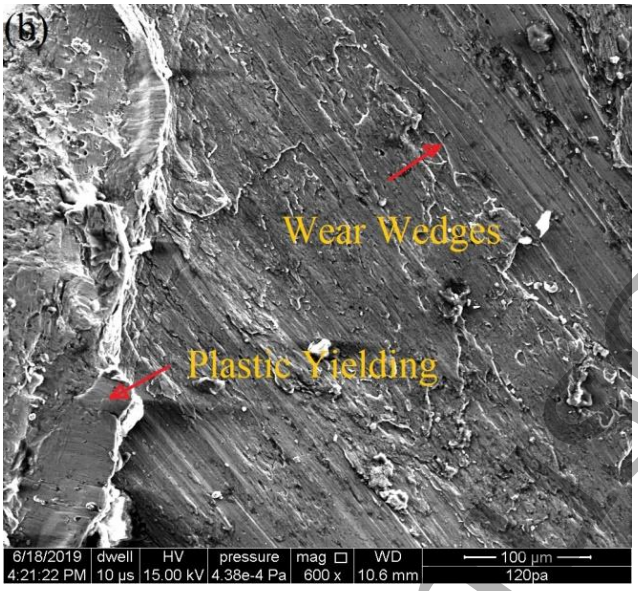


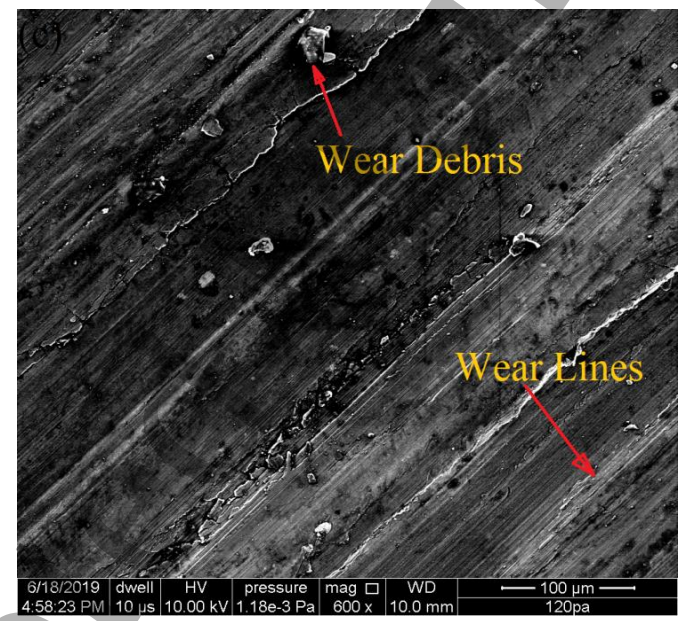
Figure 11: Coefficient of friction variation for the test specimens

For a better understanding of the wear behavior scanning electron microscope (SEM) images were obtained at 600X magnification and 15kV (Figure 12 (a-d). There were traces of wear debris on the surface but the voids were not observed. The wear area had heat-affected cliff edges. Plastic yielding was reported for sample 1, sample 2 and sample 3. However, the AA 6063 (sample 1) was the most affected with the plastic yielding. No visible yielding was observed for sample 4.



Accepted Manuscript





Accepted Manuscript

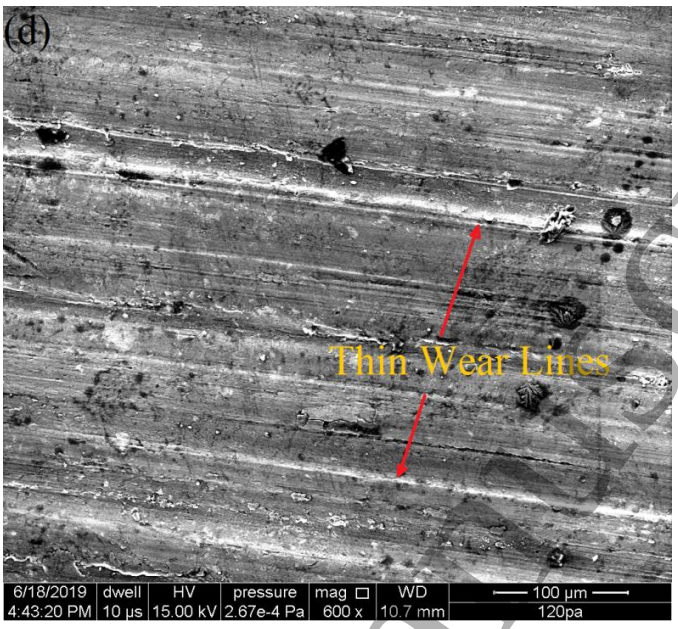


Figure 12: SEM images for (a) Al 6063 (b) composite prepared with marble dust heat to room temperature (c) composite prepared with marble dust heat to 100 ℃ temperature (d) composite prepared with marble dust heat to 200 ℃ temperature.

5. Discussion

As explained in the experimental results, it is seen that the preheating of the marble dust particles greatly influenced the mechanical properties of the aluminium metal composites. The results obtained from the microhardness measurement indicated that the marble dust particles heating to 200℃ temperature improved the microhardness of the prepared composite. However, the composite prepared with marble dust heated to 100℃ and room temperature reduced the microhardness. The low bonding between the marble dust and aluminium when preheated to lower temperature resulted in reduces microhardness. Similar trends were obtained during the tensile testing as well. The presence of moisture content in the marble dust particles at room temperature and preheated at 100℃ resulted in low bonding strength between the marble dust and aluminium particles. This resulted in poor tensile strength. With further increase in the preheating temperature the tensile strength of the processed composite gets increased. At 200℃, the residue moisture contents in the marble dust particles gets eliminate and a strong bonding between the marble dust and aluminium is obtained. Also, the completely dry marble dust preheated at 200℃ resulted with a finer microstructure. The authors did not further increase the preheating temperature beyond 200℃ as the marble dust particles were completely dried at 200℃ and heating it beyond 200℃ would not further improve the mechanical properties of the processed composites.

For the wear and friction properties, the presence of marble dust particles in the aluminium matrix improved the tribological properties. The uniform distribution of the marble dust particles provides the necessary dry lubrication between the testing surfaces which reduces the wear and friction coefficient. The scanning electron microscope images also indicated thin wear lines for the composite processed with 200℃ preheated marble dust particles.

Accepted Manuscript

6. Conclusions

In this work, the effects of preheated marble dust particles on the performance of aluminium marble dust composites was evaluated. Three aluminium marble dust composites were fabricated with the help of friction stir processing technique. The marble dust particles were preheated to a temperature of 100, 200 ℃ and reinforced in the Al 6063 material. A sample with marble dust particles heated to room temperature was also fabricated to compare the effects of preheating of marble dust on the tensile strength, microhardness, specific wear rate and coefficient of friction of the aluminium marble dust composite.

- The morphological analysis showed a finer grain size for friction stir processed aluminium marble dust composites. The preheating of the marble dust particles helped in achieving finer grain size. The composite with 200 ℃ preheated marble dust particles exhibited finer grain sizes as compared to the composite with marble dust particles heated to room temperature. The presence of marble dust was evident in friction stir processed composites. The X-ray diffraction study revealed the presence of Ca, Si, O, Al, and Mg in the processed composites.

- The micro-hardness value for the 200 ℃ preheated marble dust particles composite was higher as compared to microhardness of composite with marble dust particles heated to room temperature. There was about 29 % improvement in the microhardness of 200 ℃ preheated marble dust composite as compared to room temperature heated marble dust composite.

- The preheating of the marble dust particles also improved the tensile strength. There was a 52 % improvement in the tensile strength for the 200 ℃ preheated marble dust composite as compared to room temperature heated marble dust composite.
- The tribological properties of the fabricated composite also get improved with the preheating of the marble dust. The composite with 200 ℃ preheated marble dust resulted in a 23 % lower coefficient of friction value and 25 % low specific wear rate as compared to room temperature heated marble dust composite. The SEM images also showed that the voids, cracks, cliff edges and wear debris were lesser in case of FSP products.

References

- Alaneme, K. K., Bodunrin, M. O., & Awe, A. A. 2018. Microstructure, mechanical and fracture properties of groundnut shell ash and silicon carbide dispersion strengthened aluminium matrix composites. Journal of King Saud University - Engineering Sciences, 30(1), 96–103

- Aliabdo AA, Abd Elmoaty AEM, Auda EM. 2014 Re-use of waste marble dust in the production of cement and concrete. Constr Build Mater, 50, 28–41.

Accepted Manuscript

- Alajmi Mahdi and Shalwan Abdullah, 2015 Correlation between Mechanical Properties with Specific Wear Rate and the Coefficient of Friction of Graphite/Epoxy Composites, Materials, Vol. 8, pp. 4162-4175; doi:10.3390/ma8074162

- Apasi, D. S. Yawas, S. Abdulkwereem, and M. Y. Kolawole, 2016 Improving Mechanical Properties of Aluminium alloy through addition of Coconut Shell-ash, Journal of Science and Technology, vol. 36, no. 3, pp. 34-43.

- Aruntaş HY, Gürü M, Dayi M, Tekin I. 2010 Utilization of waste marble dust as an additive in cement production. Material and Design, 31, 4039–42.

- Atuanya U., Ibhadode A. O. A., and Dagwa I. M., 2012 Effects of breadfruit seed hull ash on the microstructure and properties of Al-SiFe alloy/breadfruit seed hull ash particulate composites, Results Phys., vol. 2, pp. 142149.

- Bilgin N, Yeprem HA, Arslan S, Bilgin A, Günay E, Maroglu M. 2012 Use of waste marble powder in brick industry Constr Build Mater 29 449–57.

- Binici Hanifi, Kaplan Hasan, Yilmaz Salih. 2007 Influence of marble and limestone dusts as additives on some mechanical properties of concrete. Sci Res Essay 2(9) 372–379

- Chaithanyasai, A., Vakchore, P. R., & Umasankar, V. 2014. The micro structural and mechanical property study of effects of EGG SHELL particles on the Aluminum 6061. Procedia Engineering, 97, 961–967.

- Demirel B. 2010The effect of the using waste marble dust as fine sand on the mechanical properties of the concrete. Int J Phys Sci 5 1372–80.

- Dinaharan, I., Balakrwashnan, M., Raja, J. D., & Akinlabi, E. T. 2019. Microstructural characterization and tensile behavior of friction stir processed AA6061 / Al 2 Cu cast aluminum matrix composites. Journal of Alloys and Compounds, 781, 270–279.

- El-Sherbiny S, El-Sheikh SM, Barhoum A. 2015Preparation and modification of nano calcium carbonate filler from waste marble dust and commercial limestone for papermaking wet end application. Powder Technol 279 290–300.

Accepted Manuscript

- El-Sherbiny S, El-Sheikh SM, Barhoum A. 2015Preparation and modification of nano calcium carbonate filler from waste marble dust and commercial limestone for papermaking wet end application. Powder Technol 279 290–300.
- Gürü M, Tekeli S, Akin E. 2007 Manufacturing of polymer matrix composite material using marble dust and fly ash. Key Eng Mater 336-338 1353–6.
- Hassan B. and Aigbodion V. S. 2015, Effects of eggshell on the microstructures and properties of Al Cu Mg/eggshell particulate composites, J. King Saud Univ. - Eng. Sci., 27(1)4956,.
- Kumar N., Saravanan R, Sellamuthu R. 2018Effect of Marble Dust on Microstructure and Mechanical properties of Al-Cu-Ni / Marble Dust Particles Composites . Int J Pure Appl Math 118: 1–12.
- Lakshmi, K. Subbaiah,J. Sarojini and Shabana 2017Study of Mechanical properties and Wear behaviour of Sugarcane ash reinforced Aluminium composite, International Journal of Mechanical Engineering and Technolog, vol. 8, no. 6, pp. 597606.
- Marras G, Careddu N, Internicola C, Siotto G. 2010 Recovery and Reuse of Marble Powder By-Product. Glob Stone Congr, 1–5.
- Mehta, K. M., & Badheka, V. J. 2019. Wear behavior of boron-carbide reinforced aluminum S.C fabricated by Friction Stir Processing. Wear, 426–427, 975–980.
- Mishra, S., Biswas, S. Satapathy, A. 2015 A study on processing, characterization and erosion wear behaviour of silicon carbide particle filled ZA-27 metal matrix composites. Materials and Design, 55, 958–965.
- Patel, S., Rana, R. S., & Singh, S. K. 2017. Study on mechanical properties of environment friendly Aluminium E-waste Composite with Fly ash and E-glass fiber. Materials Today: Proceedings, 4(2), 3441–3450.
- Prabhu, M. S., Perumal, A. E., Arulvel, S., & Wassac, R. F. 2019. Friction and Wear Measurements of Friction Stir Processed Aluminium Alloy 6082/CaCO3 Composite. Measurement, 142, 10-20.

Accepted Manuscript

- Prakash K.S., Moorthy R.S., Gopal P.M., Kavimani V. 2016, Effect of reinforcement, compact pressure and hard ceramic coating on aluminium rock dust composite performance, Int. J. Refract Metal Hard Mater. 54 223-229.
- Prasad Dora Siva and C. Shoba, 2015, Experimental evaluation onto the damping behavior of Al / SiC / RHA hybrid composites, Journal of Material Research and Technology, 5(2),123-130,.
- Prasat V., Subramanian R., Radhika N., and Anandavel B. 2011, Dry sliding wear and friction studies on AlSi10Mg fly ash graphite hybrid metal matrix composites using Taguchi method, Tribology-Materials, Surfaces & Interfaces, 5(2), 211.
- Rajak SK, Aherwar A. 2016 Physical And Mechanical Behavior Of Marble (Caco3) Filled C93200 Alloy Composites For Bearing Application. International Journal of Mechanical And Production Engineering, 4 (9), 34-37.
- Rajak SK, Aherwar A, Unune DR, Mia M, Pruncu CI. 2019 Evaluation of copper-based alloy (C93200) composites reinforced with marble dust developed by stir casting under vacuum environment. Materials 12. 1554. doi:10.3390/ma12101574.
- Senthilkumar R, Prakash M, Arun N, Jeyakumar AA. 2019The effect of the number of passes in friction stir processing of aluminum alloy (AA6082) and its failure analysis. Appl Surf Sci 491:420–31.
- Sharma, V. K., Singh, R. C., Chaudhary, R (2018): Wear and friction behaviour of aluminium metal composite reinforced with graphite particles. International Journal of Surface Science Engineering 12 419–432, 2018
- Sharma, V. K., Singh, R. C., Chaudhary, R. (2017),: Effect of fly-ash particles with aluminium melt on the wear of aluminium metal matrix composites. Eng. Sci. Technol. an Int. J. 20 1318–1323.
- Sharma, V., Singh, R. and Chaudhary(2018), R. Experimental Study of Tribological Behavior of Casted Aluminium-Bronze. Materials Today: Proceedings, 5(14), 28008-28017, 2018
- Sharma, V., Singh, R., Chaudhary, R., Saxena, M. and Anand, (2019) M. Effects of flyash addition on the dry sliding tribological behavior of aluminum composites. Materials Research Express, 6(8), 0865f4, 2019

Accepted Manuscript

- Singh, R., Chaudhary, R. and Sharma, V2019. Fabrication and sliding wear behavior of some lead-free bearing materials. Materials Research Express, 6(6), p.066533,

Accepted Manuscript

Statistical Analysis for Parametric Optimization of Gas Turbine-Steam Turbine Combined Power Cycle with Different Natural Gas Combustion

Kaushalendra Kumar Dubey, R.S. Mishra

Abstract: ***The combined Gas Turbine (GT) and Steam Turbine (ST) power generation techniques plays important role in power production. The coal fired thermal power plants have maximum possible efficiency about 34-36% with all arrangements of reheating and regeneration thermal systems. The integration of GT power generation system with old coal based thermal power plant helps to re-powering of plant and utilization of exhaust energy of GT unit through heat recovery steam generator (HRSG) of ST plant. The proposed title of research deals two steps of analysis. Energy analysis is adopted as first step for performance and energy destruction evaluation purpose where as Multi Linear regression (MLR) method is introduced as second method for parametric optimization. The four natural gases have been considered in this analysis and investigate the suitable fuel gas performance as per operating condition of GT plant such as gas turbine inlet turbine temperature (GTIT), compression ratio (CR) and mass flow of gases. The result of this paper concluded as maximum exergy loss found in combustion chamber of GT system and exhaust flow system of ST system in terms of 41% and 8% respectively. The combined and exergetic efficiency of plant are estimated to be 41% and 38.5% respectively. In present statistical model 4 levels and 3 factors (Pressure ratio, operating temperature and type of fuel gases) have been considered. And overall efficiency, gas turbine efficiency, heat loss in GT plant, Exergy destruction in thermal utilities like Compressor, combustion chamber and gas turbine are investigated. The statistical modeling concluded that the comparative results of actual and predicted results at different compression ratio of combustible gases which is affects the overall performance of combined GT-ST plant. This study helps to justify possible efficiency improvement with identifying the irreversibilities of plant utilities.***

Keywords: Exergy Analysis, Irreversibility, GT-ST combined cycle, Statistical MLR Method.

I. INTRODUCTION

The precious quantity of heat is lost during the combining of two different power generation system which will affect the plant performance. The First Law of Thermodynamic is defining the energetic criteria of system. It is incapable to give the information about thermal deficiency due to irreversibilities in process. The estimation of real performance of thermal system and quality of energy are possible by the exergy analysis. The entropy generation approach established the energy-exergy analysis for the different thermal utilities of thermal power plant and statistical methodology helps to compare the performance parameters like power output, gas turbine thermal efficiency, fuel consumption and exergetic efficiencies at different combustible gases which are utilized in combustion process. The MLR optimization method is used for the parametric analysis in present work for findings of possible overall efficiency, heat loss, and gas turbine performance with the combination of operating temperature, pressure ratio and different fuel gases. The main objective of energy-exergy modeling is to identify losses in components and true performance of plant. The several researches have been conducted on performance analysis of power plant and parametric optimization by using statistical analysis tool. The authors conducted exergy destruction analysis of gas turbine based cogeneration thermal plant and observe the effect of compression ratio, steam pressure, turbine inlet temperature, refrigeration temperature, etc. The another research is analyzed the exergy performance of ejector-absorption refrigeration system which is efficient by 8% in the application of wastage heat combined power system with refrigeration effect [1-3].Sapele-Nigeria based low power generation category of steam power plant of 75Mega Watt (MW) is computed maximum exergy destruction as 87.3% in boiler, and plant exergy efficiency 11.03%.The higher energy loss in boiler signifies the treatment of boiler utilities. The exergonomic concept for different-different power generation techniques is reviewed. Authors addressed the boiler and condenser component have major exergy destruction in case of rankine power cycle, whereas combustion chamber has major exergy destruction in case of gas-turbine powers [5-6,8].Author explain about the exergy, exergy is the maximum rate of work, which is the theoretical limitation of system and undoubtedly shows that no real system can conserve exergy, it can be recovered only [7,22].The exergetic performance of the re-powered rankine cycle with gas turbine concluded that combustion chamber and boiler have more irreversibility in GT & ST plant respectively. Researcher developed a computational model for the parametric investigation. Their analysis involved the effect of compression ratio of used combustion gas, turbine inlet temperature, boiler pressure on the energetic and

Revised Manuscript Received on November 15, 2019
Kaushalendra Kumar Dubey, Assistant Professor., Dept of Mechanical Engineering, Galgotias University, Greater Noida, India.
R.S. Mishra, Dean Outreach & Extension Activities, Delhi Technological University, Delhi India.

exergetic performance [9].The energy-exergy modeling the performance in terms of energy-exergy analysis of several thermal systems like ammonia–sodium thiocyanate absorption system, organic rankine cycle and combined cooling power generation plant (CCPP), and it provide the opportunity to improvement and identify the location of losses and losses mainly occurs during the operation and it remark as exergy destruction. The performance assessment of energy and exergy analysis of various configurations of the CCPP plant for waste heat recovery through HRSG and equivalent payback period also estimated [4,10-11].The physical systems simulation and quantitative modeling is mainly dependent on the mathematical models for accuracy and optimized design with multitude of parameters, furthermore, statistical mathematics implemented for error estimation, parametric comparison and optimization for both actual and model data examination.. For modeling and analysis generally traditional mathematical approaches have been incorporated, but complex problems can be solve by soft computing, fuzzy logic and neural networks approaches [13].The another type of plant data analysis is deals by statistical modeling techniques like regression model, least square method, Maximum Likelihood Estimator (MLE) method, Auto regression integrated moving average model (ARIMA),Multiple Linear Regression (MLR). Artificial Neural Network (ANN), Response Surface Method (RMS),Multivariate Regression (MVR), etc for error identification, parametric comparison and complex problem solving [14-19].The parametric analysis of geothermal power plant is estimated by using multiple linear regression method, and ANOVA-I&II mathematical tools have been conducted for organic rankine cycle performance analysis and compared with R-analysis tool for error estimation and parametric comparison in recent research work[20-21].

II. BRIEF OF COMBINED GT-ST POWER GENERATION SYSTEM

The proposed plant is integrated with Gas turbine and steam turbine with HRSG system. The exhaust temperature from gas turbine has potential to run the steam turbine plant through HRSG. The steam power plant thermal efficiency is fixed with maximum possible efficiency 36% because the gas turbine power cycle is works as toping cycle and the variation in GT plant will affect the overall performance of plant. The layout of proposed plant is made in AUTOCAD which is shown in fig-1; operational conditions and combustible gas properties have been listed in Table-1&2.

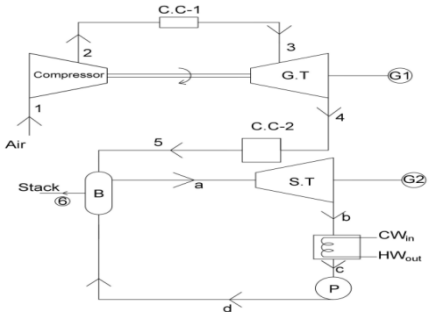


Fig.1. Thermodynamic schematic of combined GT-ST plant

schematic The combined GT-ST plant is integrated with GT power generation and ST power generation, the following process are given below as per the fig-1 mentioned plant.

Process: 1-2- Isentropic compression of atmospheric air in GT plant.
Process: 2-3-Combustion of gas in combustion chamber-1 of GT plant.
Process: 3-4-Isentropic flue gas expansion in gas-turbine of GT plant.
Process: 4-5-Combustion of flue gas in combustion chamber-2 of GT plant.
Process: 5-6-Stack flow out.
Process: 6-a-Steam generation in HRSG of ST plant by utilizing exhaust heat from GT plant.
Process: a-b-Isentropic steam expansion and steam-turbine work perform of ST plant.
Process: b-c- steam condensation and heat rejection of ST plant.
Process: c-d-pumping of condensed water & makeup water and supply to HRSG of ST plant.

III. PLANT OPERATION CONDITION AND COMBUSTIBLE NATURAL GAS PROPERTIES

Table 1. Operating Condition of GT-ST plant

Plant Components/Parameters	Unit & Value	Plant Components/Parameters	Unit And Value
Compressor Inlet condition (P1,T1)	1bar,25 ^{0}C	Specific heat ratio of air ($\gamma_{_air}$)	1.4
Compressor Pressure Ratio (P2/P1)	8	Steam condition at inlet of Steam turbine	40bar,425 ^{0}C
Gas Turbine Inlet gas temperature (T3)	900 ^{0}C	Condenser pressure	0.04 bar
Pressure drop in combustion chamber	3%	Feed water temp to HRSG	170.4 ^{0}C
Compressor Efficiency (η_c)- manufacturer efficiency	88%	ST Efficiency (η_{ST})	82%

GT Efficiency (η_{GT})-manufacturer efficiency	88%	Pressure Drop of gas in the HRSG	5kPa
Calorific Value of liquid Ethane as fuel (CV_f)	44.43MJ/Kg	Steam Flow rate= w_s	105 TPH
Specific heat of air ($Cp_{_air}$)	1.006 KJ/KgK	GT outlet Pressure (P4)	1.05 bar
Specific heat of gas ($Cp_{_gas}$)	1.148 KJ/KgK	Condenser Inlet staem or ST outlet flow	105TPH or 30Kg/s

IV. ENERGY ANALYSIS AND STATISTICAL MODELING OF COMBINED GT-ST POWER PLANT

The governing equation of analysis is based on energy-mass conservation and entropy generation principle of thermodynamics. The MLR method is used for parametric optimization for the GT-ST plant performance. Mass and energy balance equations has been applied in all thermal utilities. In order to simplify the analysis, some fundamental assumptions and proposed analysis is adopted from PK Nag combined GT-ST model [12]:

1. The thermodynamic process is assumed steady flow with consideration of control volume (CV) system.
2. The isentropic performance is considered for compressor, pumps and both GT& ST turbines.

4.1 Energy Analysis

Applying mass-energy and 1st law energy equation in all utility of GT plant. The thermodynamic relation and equations for power plant components are as follows

Compressor

The compressor work is function of operating condition of air intake; the compressor outlet temperature can be expressed as following thermodynamic relations-

$$\frac{T2}{T1} = \left(\frac{P2}{P1}\right)^{\frac{(\gamma-1)}{(\gamma * \eta c)}} \quad (1)$$

$$W_comp = m_air \times Cp_air (T2-T1) \quad (1a)$$

Combustion Chamber

The pressure drop (ΔPcc) across the C.C as follows

$$P3/P2 = (1- \Delta Pcc) \quad (2)$$

Assume flow rate of flue gas is 1Kg/s and that of fuel f Kg/s

The mass flow rate of air = (1-f) Kg/s (3)

Therefore $f \times CV_f = mf \times Cp_{_gas} (T3-T4) - (1-f) Cp_air (T2-T0)$ (4)

$$\text{Air Fuel Ratio} = \frac{A}{F} = \frac{(1-f)}{f} \quad (5)$$

Gas Turbine

The GT outlet temperature is estimated by isentropic efficiency (η_{GT})

$$\frac{T3}{T4} = \left(\frac{P3}{P4}\right)^{\frac{(\gamma-1)(\eta GT)}{(\gamma)}} \quad (6)$$

Heat Recovery Steam Generator (HRSG)

The flue gas temperature at inlet of HRSG is calculated by

Let the pinch point temp differences (T5-Tf) be 30 ^{0}C (pinch point at exit of GT and inlet of ST)

From Mass and Energy balance between GT and ST

$$m_gas \times Cp_gas (T4-T5) = m_steam (h_a-h_f) \quad (7)$$

And air flow rate entering the compressor = m_air = (1-f) m_gas (8)

Fuel mass flow rate $m_fuel = w_f = f \times m_gas$ (9)

Than energy interaction between HRSG and stack flow, and temperature of stack flue gas as follows

$$1.14 (482-T6) = 0.106 (3272-721.1) \quad (9a)$$

Stack flow Temp= T6

Power Output of GT-ST Plant-

The combined GT-ST plant power output is achieved by the net work done by gas turbine and steam turbine both, total work of GT& ST plant is give by

$$W_total = W_{_ST\ plant} + W_{_GTplant}$$

Energy equation for steam turbine work in terms of enthalpy across inlet and outlet of ST as follows

$$W_{_ST} = w_s (ha - h_{bs}) \times \eta_st \quad (10)$$

And Enthalpy at condenser line = $h_{bs} = h_{f_b} + x_bs \times h_{fg_b}$ (11)

Table 2. Properties of combustible gas for GT plant

S. No	Name Of natural gas	Chemical Formula	Heat Value (MJ/KG)	Cp/Cv	γ	Combustion Reaction (CaHb+c*O2-D*CO2+E*H2O)	Stoichiometric Combustion A/F(A/F=C*O_mol wt/(23.2% * Fuel _mol wt))
1	Methane	CH4	55.5	2.22/1.70	1.3	CH4+2O2-CO2+2H2O	17.24
2	Ethane	C2H6	51.9	1.53/1.48	1.03	C2H6+7/2O2-2CO2+3H2O	16.09
3	Propane	C3H8	50.4	1.67/1.48	1.12	C3H8+5O2-3CO2+4H2O	15.6
4	Butane	C4H10	49.1	1.67/1.53	1.091	C4H10+13/2O2-4CO2+5H2O	15.45
5	Pentene	C5H12	48.6	1.66/1.52	1.092	C5H12+8O2-5CO2+6H2O	15.32

Work done by GT plant is depend on the difference between gas turbine work and compressor work consumption, the combine equation for $W_{_GT}$ and $W_{_COMP}$ as follows

$$W_{_GT}\ plant = W_{_GT} - W_{_COMP} \quad (12)$$

$$W_{_GT}\ plant = m_gas \times Cp_gas (T3-T4) - m_air \times Cp_air (T2-T1) \quad (13)$$

When two plants are combined, there is always some heat loss. If heat rejected by GT plant as topping cycle is absorbed by ST plant as bottoming cycle, The Lost heat as coefficient in the exhaust stack

$$X_L = \left(\frac{wg \times Cpg\,(T6-T1)}{wf \times CVf}\right) \quad (14)$$

For the overall thermal efficiency with considering heat lost between topping and bottoming cycle. The overall efficiency of plant as follows

$$\eta_{overall\ plant} = \eta_{STplant} + \eta_{GTplant} - \eta_{STplant} \times \eta_{GTplant} - \eta_{STplant} \times X_L \quad (15)$$

The GT-ST combined plant efficiency is depend on work done by both turbines and heat supplied through combustion chamber and HRSG respectively in GT and ST plant. The equations for efficiencies are follows

$$\eta_{STplant} = (ha-hb)/(ha-he) \quad (16)$$

put all values of enthalpy from equ. No 10-11-12-13

$$\eta_{GTplant} = (W_{GT})/(w_f \times CV_f) \quad (17)$$

From Equ No 15 $\boldsymbol{\eta_{overall}}$

4.2 Exergy Destruction Analysis

Assume exergy flux $\psi = (\Delta G_0/\Delta H_0) = 1.0401 + 0.1728\ (h/c)$

(18)

Where $\Delta H_0 = w_f \times (CV)_0$

Compressor-

Rate of energy dissipation in compressor (I_{comp}) $= w_a T_0 \left(Cp_a \ln \frac{T2}{T1} - R_a \ln \frac{P2}{P1}\right)$ (19)

But $R_a = C_p \left(\frac{\gamma-1}{\gamma}\right)$

Combustion Chamber-

The irreversibility in CC is estimated by the energy balance between product, reactants, air and used fuel respectively.

$$IccT_0\left[wg\left\{Cpg\ ln\frac{T3}{To} - Rg\ ln\frac{P3}{Po}\right\} - \left\{wgCpg\ ln\frac{T2}{T0} - waRa\ ln\frac{P2}{Po}\right\} + \Delta S\right] \quad (20)$$

Gas Turbine-

Rate of energy lost or work lost in GT = $I_{GT} = w_g T_0 \left[Cpg \ln \frac{T4}{T3} - Rg \ln \frac{p4}{p3}\right]$ (21)

HRSG-

Rate of energy lost in heat recovery steam generator-$I_{HRSG} = T_0[ws(sa - se) + Cpg\ ln\frac{T6}{T4} - Rg\ ln\ \frac{P6}{P4}]$ (22)

Steam Turbine

Rate of energy or work lost in the steam turbine= I_{ST}

$I_{ST} = T_0\ w_s\ (s_b - s_a)$ (23)

Exergy Lost due to exhaust flue gas

$I_{EXHFLUEGAS} = \int_{T0}^{T6}\left(1 - \frac{T0}{T}\right)dQ = w_g \times Cpg\ [(T6 - T0) - T0\ ln\frac{T6}{T0}]$ (24)

Exergetic Efficiency=$\eta_{EX} = \frac{Total\ Output}{Exergy\ Input}$ (25)

4.3 Statistical Analysis-MLR Regression Method

The MLR method of statistical model is used in this analysis for overall and gas turbine both. The four different gases have been used in combustion chamber at different pressure ratio and operating temperature ranges.

This helps to identify the suitable combination of input factors for optimized result of plant operation.

The mathematical expression for expected value of performance parameters as follows which give the optimized result in all set of available factors and level of operating conditions.

Table 3. Factors and Level of combined GT-ST power plant

Factors (Input operational factor)	LEVEL 1	2	3	4	
A-Compressor Pressure in Bar	6	7	8	9	
B-Operating Temperature at inlet of GT in Degree C	600	700	800	900	
C-Fuel Gases	Butane	Methane	Ethane	Octane	

Above mentioned factors and its level analyzed by L16 model of MLR analysis.

MLR Regression Analysis-

Regression Equation

Overall Effi (%) = 36.62 + 0.428* Pressure +2.982*Temperature- 0.588 HC

(Neglecting effect of fuel gas composition, only compression pressure and GTIT consider)

where, Overall efficiency represents the gas turbines inlet temperature and gas turbines cycle pressure .

Coefficients

Term	Coeff	SE Coeff	T-Value	P-Value	VIF
Constant	36.62	2.48	14.79	0.000	
Pressure(Bar)	0.428	0.553	0.77	0.454	1.00
Temperature (Degree Celsius)	2.982	0.553	5.39	0.000	1.00
HC	-0.588	0.553	-1.06	0.309	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.47505	71.93%%	64.92%%	51.99%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	188.405	62.802	10.25	0.001
Pressure(Bar)	1	3.668	3.668	0.60	0.454
Temperature (Degree Celsius)	1	177.817	177.817	29.03	0.000
HC	1	6.921	6.921	1.13	0.309
Error	12	73.510	6.126		
Total	15	261.916			

V. RESULTS AND DISCUSSION

The general output and exergy of components of proposed

Table 4. Observation Tables of GT-ST Plant performance (operated with octane gas)

Performance Parameters	Parametric Values	Performance Parameters	Parametric Values	Performance Parameters	Parametric Values
$W_{_STplant}$	29MW	Gas Turbine Exergy Destruction	5646 KW	Exhaust Flue Gases Exergy Destruction	17760 KW
$W_{_GT\ plant}$	53MW	HRSG Exergy Destruction	7583 KW	Total I_{LOSS} Or Destruction = $\sum Iloss$ =	132945 KW
$\eta_{STplant}$ & $\eta_{GTplant}$ & $\eta_{overall\ plant}$	38% & 27% & 41.7%	Steam Turbine Exergy Destruction	6412 KW	Exergy Input (ΔG_0)	212810 kW
Heat lost during combining of GT-ST cycle =X_L	35.40%	Compressor Exergy Destruction	6613 KW	Total Output	82,000KW

m_gas / m_air /m_fuel	275.4Kg /s /271.3kg /s /4.46 Kg	Combustion chamber Exergy Destruction	88661 KW	Overall Exergetic Efficiency= η_{EX}	**38.50%**

GT-ST combined plant are given in table-4. The present result is considering octane gas as a fuel gas in gas turbine plant with the all permissible operational condition for plant working.The performance of plant is observed by using energy analysis of different units of proposed plant. Efficiencies, Heat loss, mass flow of air & gas and exergy destruction are major performance indicators which are provided in table - 4. 38%,27% and 41.7 % of ST plant efficiency, GT plant efficiency and combined GT-ST plant efficiency are estimated respectively with the 35.4% of heat loss during the GT and ST cycle combination. The combustion chamber of GT plant has higher energy loss with 41.6%, where as GT is utilize maximum energy of input. The overall efficiency of plant is estimated as 41.7%, but the actual plant performance is calculated in terms of exergetic efficiency as 38.5%.

5.1 Effect of performance parameters:

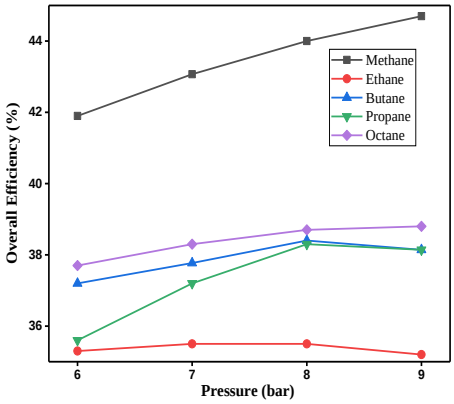


Fig 2(a).Effect of inlet pressure of GT on η_overall

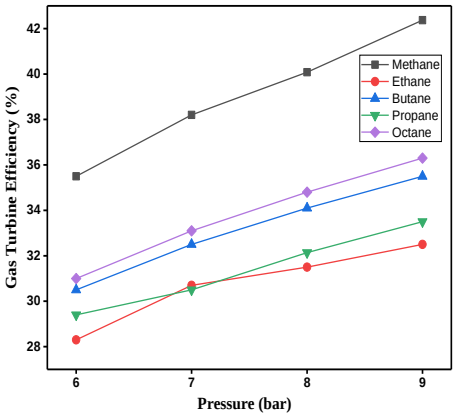


Fig 2(b). Effect of inlet pressure of GT on η_GT

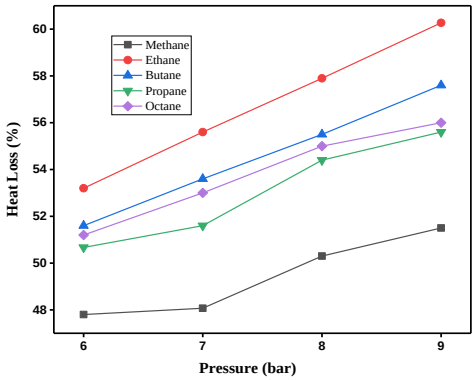


Fig 2(c).Effect of inlet pressure of GT on Heat loss between GT&ST

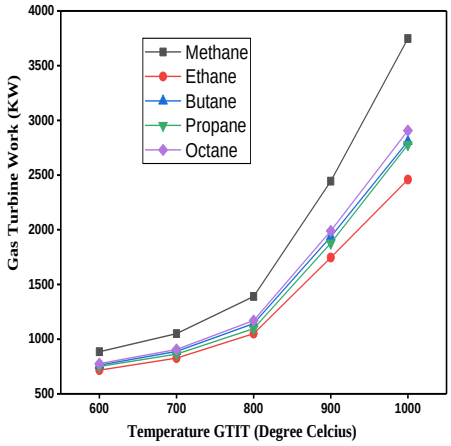


Fig 2(d). Effect of gas turbine inlet temperature in GT Work

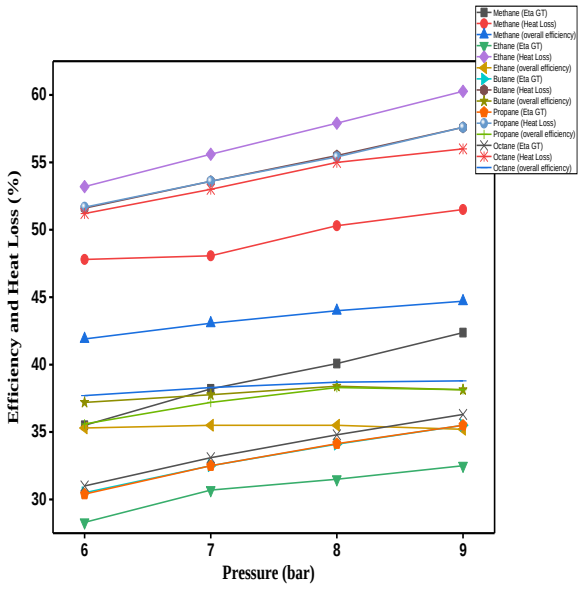


Fig 2(e) Effect of inlet pressure of GT in overall performance

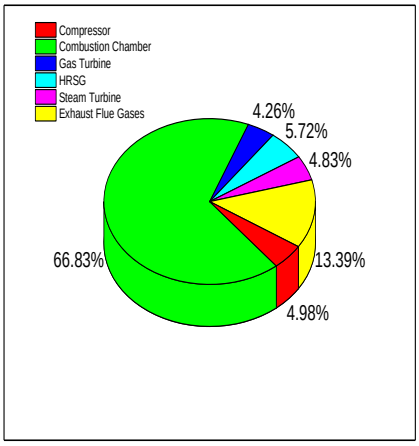


Fig 2(f). Exergy Destruction of Major component of combined GT-ST Plant

The more compression ratio enhances the GTIT and higher temperature of GT inlet gives higher thermal efficiency. Fig 2(a), 2(b), 2(c), 2(d) & 2(e) are showing the effect of compression ratio in overall efficiency of combined GT-ST plant with five different gas combustion in GT plant. The maximum GTIT taken 1000 ^{0}C, the metallurgical criteria of turbine material is main factor for limiting value of GTIT. Methane gas have higher efficiency with high CR and GTIT values, fig 2(c) clearly indicate minimum heat loss also, where as other gases have valuable output with all ranges of CR values. Overall performance of plant with all composition of natural gases is mentioned in fig 2(e).

The Total exergy loss of plant is estimated by 133MW with the 213 MW of input energy through CC and HRSG heat addition as mentioned in table-4, the component wise irreversibility is explained by fig-2(f) The maximum exergy lost 66.83% is found in CC of GT system and exhaust flue gas unit of ST system about 13.39%. Steam turbine also has more energy loss as compare to gas turbine.

5.2 Statistical modeling Result

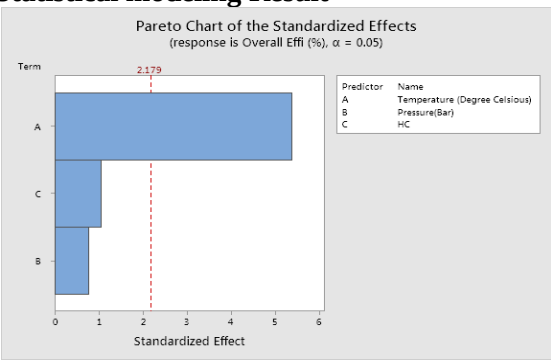


Fig-3(a). Pareto Chart of overall efficiency

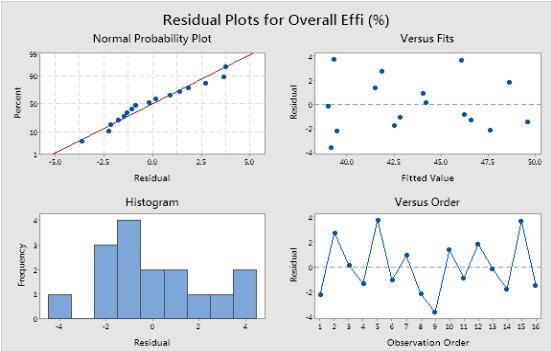


Fig-3(b) Regression Fit chart for overall efficiency

Table-5 L16 Regression model for Actual and Predicted result comparison

Compressor Pressure(Bar)	Temperature – GTIT (Degree Celsius)	Natural Gas	Actual Overall Efficiency (%)	Predicted Overall Efficiency (%) By DOE	Residual	% Deviation
6	600	Butane	37.2	40.03	-2.83	-7.607526882
6	700	Methane	44.6	43.012	1.588	3.560538117
6	800	Ethane	44.4	45.994	-1.594	-3.590090009
6	900	Octane	45.3	46.976	-1.676	-3.699779249
7	600	Methane	43.07	40.458	2.612	6.064546088
7	700	Butane	41.8	43.44	-1.64	-3.923444976
7	800	Octane	45	46.422	-1.422	-3.16
7	900	Ethane	45.5	49.404	-3.904	-8.580219978
8	600	Ethane	35.5	40.886	-5.386	-15.17183099
8	700	Octane	42.9	43.886	-0.986	-2.298368298
8	800	Butane	45.39	46.85	-1.46	-3.216567526
8	900	Methane	50.5	49.482	1.018	2.015841584
9	600	Octane	38.8	41.314	-2.514	-6.479381443
9	700	Ethane	40.78	44.296	-3.516	-8.621873467
9	800	Methane	49.8	47.278	2.522	5.064257028
9	900	Butane	48.19	50.26	-2.07	-4.295496991

% of Deviation = [(actual value – predicted value)/ actual value] × 100%

The MLR model summery measure the overall efficiency of the plant. The adjusted R-squared value of 64.92% and the Predict R-squared value of 51.99 % are fairly acceptable for this analysis. Furthermore, a value of R^2 on a higher side indicates that the results are close to the real data values. A comparison between the actual data and the predicted one is highlighted in Table 5. A set of data based on 16 runs were collected as per the operating gas turbine pressure, GTIT and combustible natural gases. The table was constructed which included the predicted responses from MLR analysis. The values seem to coincide with the actual data which is considered good. It also shows the residual by response surface and the percentage variation versus the run number in fig-4. A normal plot of residuals is shown in Figure-3(b). The graphs highlight one key fact, that is, the value of R^2 is 71.93%, shows minimum irregularity. The linear model is statistically adequate. It can be seen that the best-fit line resembles the (Y=X) line when plotted on a data set of 16 points. This indicates that the predicted

values and the real data are in close proximity to each other. The variation found in actual results from the predicted ones was between 2.612 and -5.386 showing only a minor dissimilarity. The pareto graph fig-3(a) well explained the GTIT is main factor for overall efficiency with the set of compressor pressure and fuel gases.

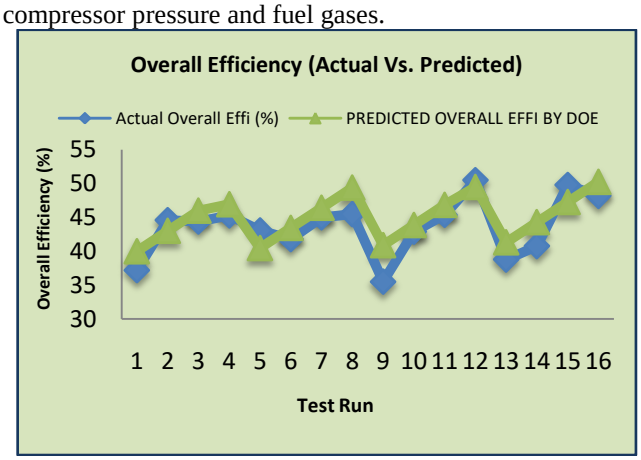


Fig-4 Overall efficiency (Actual Vs Predicted by DOE)

VI. CONCLUSIONS

The present analysis is carried energy and exergy approach of thermodynamics for actual performance where as MLR regression modeling is applied for comparison between actual and predicted experiments results. The highlights of this analysis have mentioned below-

1-The performance of combined GT-ST plant is influenced by gas turbine performance and as well as effected by heat loss. Steam turbine plant efficiency is fixed by 38% which is maximum possible efficiency.

2-The assessment of exergy destruction is important for actual plant performance. The determination of exergy loss estimation of various components helps to identify the components for control in heat loss or re-design of identified parts. Combustion chamber and heat recovery steam generator have major exergy loss in GT and ST plant respectively. Steam turbine exergy destruction is more than gas turbine destruction due to pressure reduction and steam condensation during the expansion of steam. It indicates the temperature at the inlet of steam turbine must be high as per materialistic temperature limit of turbine material. Superheated steam is required for the achievable turbine performance.

4. The MLR statistical highlight one key fact, that is, the linear model is statistically acceptable due to the best-fit line when plotted on a data set of 16 test runs. This indicates that the predicted values and the real data are in close proximity to each other. The variation found in actual results from the predicted in terms of average absolute residuals and average value of % deviation values are -1.328 and -3.371 respectively for the case of overall efficiency analysis of plant. It was due to cyclic variations that accounted for this minor difference between the real and the predicted data.

5. The overall plant output greatly affected by the heat loss. The maximum possible GT efficiency is achieved as 46% at higher CR value with methane gas combustion as shown in fig 2(a) & 2(b) .The heat supply through combustion chamber is converted in to GT work done and GT cycle heat rejection further, but rejected heat of GT cycle is utilize through HRSG of steam power plant for ST work output. If heat loss is more, than overall performance will decrease.

The both approach of analysis gives the comparative and optimized solution for plant operation at suitable operating condition. The present statistical model of DOE helps to compare the actual and predicted values for overall plant performance study, machinery optimization, components identification for controlling heat loss, best combination of factor with the large range of input operating parameters.

REFERENCES

1. A Khaliq,SC Kaushik, (2004), Thermodynamic performance evaluation of combustion gas turbine cogeneration system with reheat, Journal of applied thermal engineering, Elsevier, 24, pp1785-1795. .
2. Khaliq and S.C Kaushik "Second-law based thermodynamic analysis of brayton/rankine combined power cycle with reheat" Applied Energy, Vol 78,No.2,pp 179-197,2004.
3. [3] Abdul Khaliq, Exergy analysis of gas turbine trigeneration system for combined production of power heat and refrigeration, Int. J. Refrig. 32 (3) (2009) 534–545, http://dx.doi.org/10.1016/j.ijrefrig.2008.06.007.
4. Omendra Singh,SC Kaushik,(2013), Thermodynamic evaluation and optitimization of a Brayton-Rankine-Kalina combined triple power cycle,Journal of energy conversion and management,71,pp 32-42.
5. Christian O.Osueke, Anthony O. Onokwai and AdeyinkaO.Adeoye "Energy and exergy analysis of 75 MW steam power plant in Sapele (Nigeria) "International journal of innovative research in advanced engineering.ISSN-2349-2163,vol-2,issue-6,pp-169-179, 2015.
6. Ravindra kumar "A critical review on energy,exergy,exergonomics and economic (4-E) analysis of thermal power plants"Engineering Science and Technology, an International Journal,Elsevier,issn-2215-0986,pp 283-292, Feb 2017
7. P. Ahmadi, Marc A. Rosen, I. Dincer, Greenhouse gas emission and exergoenvironmental analyses of a trigeneration energy system, International Journal of Greenhouse Gas Control 6 (2011) 1540e1549.
8. L.G. Chen, H.J. Feng, F.R. Sun, Exergy optimisation of irreversible closed Brayton cycle combined cooling, heating and power plant, J. Energy Inst. 86 (2) (2013) 97–106, http://dx.doi.org/10.1179/1743967112z.00000000048.
9. Arvind Kumar Tiwari, M.M. Hasan, Mohd Islam, Exergy analysis of combined cycle power plant: NTPC Dadri, India, Int. J. Thermodyn. 16 (1) (2012) 36–42, http://dx.doi.org/10.5541/ijot.443.
10. Thamir K. Ibrahim, Mohammed Kamil Mohammed , Omar I. Awad, Ahmed N. Abdalla, Firdaus Basrawi, Marwah N. Mohammed, G. Najafi, Rizalman Mamat, A comprehensive review on the exergy analysis of combined cycle power plants, Renewable and Sustainable Energy Reviews 90 (2018) 835–850.
11. Jing Li, Guangtao Gao, Cagri Kutlu, Keliang Liu, Gang Pei, Yuehong Su,A novel approach to thermal storage of direct steam generation solar power systems through two-step heat discharge. Applied Energy Volume 236, February 2019, Pages 81-100.
12. P. K. Nag, "Power Plant Engineering" 3rd edition,Tata McGraw Hill Publishing India Pvt Ltd., ISBN-13: 978-07-064815-9, ISBN-10:0-07-064815-8, (2008).
13. Yang L, Entchev E. Performance prediction of a hybrid microgeneration system using Adaptive Neuro-Fuzzy Inference System (ANFIS) technique, Applied Energy. 2014,134, 197-203.
14. C.M. Douglas, A.P. Elizabeth, G.V. Geoffrey.Introduction To Linear Regression Analysis,(5th Edition), John Wiley and Sons (2011).
15. L.J. Scott, F. Jeremy.Regression Models for Categorical Dependent Variables using Stata,(4th Edition), Stata press, Texas (2014).
16. W.G. Martin, W.S. Maynard Predictive modeling of multivariable and multivariate data,J. Am. Stat. Assoc. (2012), pp. 646-653
17. S.T. Au, G.Q. Ma, R. Wang,Iterative multivariate regression Model for correlated response prediction Int. Conf. on Cyber enabled Distributed computing and Knowledge Discovery, 5(2011), pp. 55-59.
18. R.D. McKelvey, W. Zavoina.A statistical model for the analysis of ordinal level dependent variables,J. Math.Sociol., 4 (1) (1975).
19. M. Karadas, H.M. Celik, U. Serpen, M. Toksoy.Multiple regression analysis of performance parameters of a binary cycle geothermal power plantGeothermics, 54 (2015), pp. 68-75
20. U. Larsen, L. Pierobon, J. Wronski, F. Haglind.Multiple regression models for the prediction of the maximum obtainable thermal efficiency of organic Rankin cycle,Energy, 65 (1) (2014), pp. 503-510.

21. Y. Tunckaya, E. Koklukaya.Comparative prediction analysis of 600 MWe coal-fired power plant production rate using statistical and neural based models,J. Energy Inst., 88 (1) (2015), pp. 11-18.
22. Kaushalendra Kumar Dubey,R.S.Mishra, Energy-Exergy Analysis Of Combined Reheating-Regenerative Rankine Cycle Using Entropy Generation Principle, International Journal Of Recent Technology And Engineering (IJRTE),Volume-8, Issue-3, September 2019,pp.3340-44.

AUTHOR'S DETAIL

Kaushalendra Kr. Dubey is associated with is the Dept of Mechanical Engineering, Galgotias University, Greater Noida, India as assistant professor. He is a research fellow of Delhi Technological University, Delhi, India. Author qualification is Master of technology in Thermal Engineering with honors from Jamia Millia Islamia, Delhi, India..He has faculty research fellowship from Indian Institute of Technology, Delhi in 2016 in Energy and Heat Recovery Research. Author has published more than 25 research papers in repute journals and conferences and active volunteer for work on energy research and technology through industrial and academic project.

Prof. (Dr.) R.S. Mishra Ph.D-Solar Energy IIT Delhi-1986.Dean Outreach & Extension Activities, Delhi Technological University, Delhi India . He has published more than 250 research papers in repute journals & conference. Professor Mishra is dynamic personality of area of solar thermal system design-development and advanced refrigeration techniques. He is awarded by many national and International scientific societies.

water

MDPI

Article

Temporal Probability Assessment and Its Use in Landslide Susceptibility Mapping for Eastern Bhutan

Abhirup Dikshit [1], Raju Sarkar [2,3], Biswajeet Pradhan [1,4,*], Ratiranjan Jena [1], Dowchu Drukpa [5] and Abdullah M. Alamri [6]

1 Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering & IT, University of Technology Sydney, Sydney, NSW 2007, Australia; abhirupdikshit@gmail.com (A.D.); ratiranjan.jena@student.uts.edu.au (R.J.)
2 Department of Civil Engineering, Delhi Technological University, Bawana Road, Delhi 110042, India; rajusarkar@dce.ac.in
3 Department of Civil Engineering and Architecture, College of Science and Technology, Royal University of Bhutan, Rinchending 21101, Bhutan
4 Department of Energy and Mineral Resources Engineering, Sejong University, Choongmu-gwan, 209, Neungdongro Gwangjin-gu, Seoul 05006, Korea
5 Department of Geology and Mines, Ministry of Economic Affairs, Royal Govt. of Bhutan, Thimphu 11001, Bhutan; ddukpa@moea.gov.bt
6 Department of Geology & Geophysics, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia; alamri.geo@gmail.com
* Correspondence: biswajeet.pradhan@uts.edu.au

Received: 25 November 2019; Accepted: 15 January 2020; Published: 17 January 2020

Abstract: Landslides are one of the major natural disasters that Bhutan faces every year. The monsoon season in Bhutan is usually marked by heavy rainfall, which leads to multiple landslides, especially across the highways, and affects the entire transportation network of the nation. The determinations of rainfall thresholds are often used to predict the possible occurrence of landslides. A rainfall threshold was defined along Samdrup Jongkhar–Trashigang highway in eastern Bhutan using cumulated event rainfall and antecedent rainfall conditions. Threshold values were determined using the available daily rainfall and landslide data from 2014 to 2017, and validated using the 2018 dataset. The threshold determined was used to estimate temporal probability using a Poisson probability model. Finally, a landslide susceptibility map using the analytic hierarchy process was developed for the highway to identify the sections of the highway that are more susceptible to landslides. The accuracy of the model was validated using the area under the receiver operating characteristic curves. The results presented here may be regarded as a first step towards understanding of landslide hazards and development of an early warning system for a region where such studies have not previously been conducted.

Keywords: shallow landslide; landslide susceptibility; temporal probability; Bhutan

1. Introduction

Rainfall triggered landslides are one of the most devastating naturally occurring disasters across the world [1]. The global dataset of landslide hazards in the 2004–2016 period extracted from Reference [2] showed that almost 75% of the world's fatal landslides occurred in the Himalayan region. Bhutan is no exception to this, and is a part of one of the world's highly landslide-prone regions in the world [3]. The damage caused by landslides in this country has led to casualties and loss of land, affecting people's livelihoods and disrupting the transportation network, which is key to the country's economy. Most of the landslides in the Bhutan Himalayas are triggered by rainfall, especially during the monsoon period [4,5]. Therefore, it is imperative to identify the areas that could be affected by

landslides, in order to reduce the probability of damage in the future. The key to achieving this is through a detailed landslide hazard assessment that will help civic authorities to curtail landslide damage through effective land use management.

Landslide hazard may be defined as the probability of a damaging landslide in a spatial ("where") and temporal ("when") context, along with the magnitude ("how large") of the event [6,7]. Landslide susceptibility is defined as the likelihood of landslide occurrence ("where") in an area depending on local terrain conditions [8]. It may be regarded as the first step towards analyzing hazard and risk. Various spatial assessment models for landslide susceptibility have been developed [9–11]. However, compared to spatial assessment, there have been fewer attempts to carry out temporal probability assessment (studies have been conducted in Nilgiris, India [12], Hoa Binh, Vietnam [13], and Cameroon [14]). The two main techniques used to assess temporal probability for future landslide occurrences are (i) analysis of potential slope failure and (ii) statistical analysis of past landslide events [13,15]. The first technique involves evaluation of the current slope conditions and determination of the probability for future slope instability, which may be difficult to apply in large study areas [16]. Statistical analysis of past landslide events may be done directly using records of the landslides identified in the study area or, alternatively, it may be performed indirectly by using information related to recurrence of the landslide-triggering events [17]. Direct analysis requires a long time span of historical landslide data which is extremely difficult to obtain, especially in underdeveloped countries. Therefore, an indirect approach analyzing the frequency of occurrence of rainfall was used in this study to determine temporal probability. Even though this approach did not require complete multi-temporal landslide inventory data, it required determination of the relationship between rainfall and landslide incidences. After the calculation of rainfall thresholds, the landslide temporal probability was computed based on the number of times precipitation exceeded the threshold value [16]. As the frequency of rainfall-induced landslides only evaluates how often landslides might occur, it therefore needs to be integrated with spatial probability (susceptibility) and temporal probability to develop a landslide hazard map [17,18].

The prediction of landslide incidences using rainfall thresholds has been successfully carried out for various regions, including Italy [19–21], New Zealand [22], Malaysia [23], and the Himalayan arc [24–26]. The calculation of rainfall thresholds for landslide triggering can be determined using three main approaches: (i) physically based models [27], (ii) empirical rainfall threshold models [28], and (iii) statistically based models [29]. The physically based models are linked to the physical attributes of the study region and can be difficult to apply in cases of unavailability of an extensive dataset. Empirical models calculate rainfall thresholds based on past rainfall events which led to landslide incidences. The threshold is usually obtained by drawing lower-bound lines to the rainfall conditions that resulted in landslides plotted in Cartesian, semi-logarithmic, or logarithmic coordinates [30]. Statistical models use statistical tools like Bayesian inference or logistic regression to calculate thresholds [31].

In the case of Bhutan, studies to date have focused on the southwest region covering the Phuentsholing–Thimphu Highway (known as the Asian Highway), which connects the national capital Thimphu with neighboring countries. These studies have primarily focused on rainfall estimation and spatial assessment, using various techniques such as a probabilistic approach [5,32], a semi-automatic algorithm approach [26], an empirical approach [4], and machine learning models [33]. The other major highway, Samdrup Jongkhar–Trashigang (S-T), situated in the eastern part of the country, has been neglected, and a landslide study in this region is yet to be conducted. The main aim of this study was to assess landslide susceptibility utilizing temporal rainfall for the S-T highway. The two objectives in this study were (i) to estimate the temporal probability, and (ii) to estimate the landslide susceptibility using a multi-criterion decision-making approach. We addressed three major research themes in the current study: (i) determination of rainfall threshold, probability estimation of the threshold being exceeded, and landslide probability after the threshold has exceeded; (ii) susceptibility of the region with respect to landslides; and (iii) validation of the thresholds and susceptibility map. For this, the rainfall thresholds were determined based on the relation between daily rainfall and past landslide events that occurred between 2014 and 2017. The thresholds were validated using the rainfall records

of 2018. Thereafter, the exceedance probability of the threshold was calculated and the temporal probability of landslides was determined using a Poisson model. Finally, a landslide susceptibility map was developed using the analytic hierarchy process (AHP), utilizing the determined threshold values. This study was the first attempt to conduct such an elaborate study for the eastern region of Bhutan. The results from the present work can be understood as a preliminary step towards setting up an operational landslide early warning system so that damage to the transportation corridor can be reduced and human lives can be saved.

2. Study Area

The study area was the Samdrup Jongkar–Trashigang (S-T) highway, which is a 180 km stretch of road located in the eastern part of Bhutan, which covers 1880 km^2 (Figure 1). The region was selected as it connects eight districts (known as "dzongkhags" in Bhutanese) and is a major route for the people residing around the highway. The highway is critical as it is the only transportation network connecting eight dzongkhags in East Bhutan, and it is a lifeline for the people residing in these areas. The transportation corridor has a history of severe slope failures in the form of frequent landslides, rockfalls, and mudflow. The region falls in a sub-tropical zone, where heavy rainfall is frequent. The region receives its maximum rainfall during the monsoon seasons (June–September). The rainfall pattern in this region can be described as low-intensity and long-duration with occasional intermittent outbursts.

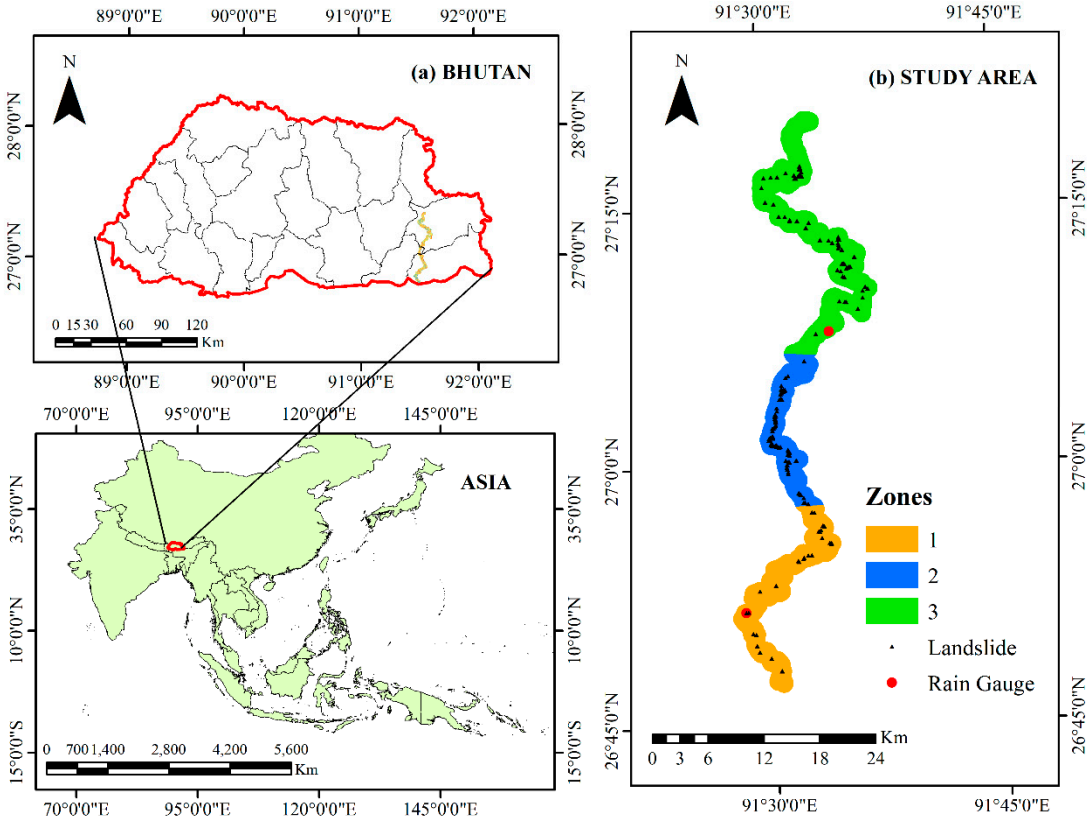


Figure 1. Location of (**a**) Bhutan; (**b**) spatial distribution of landslides in the study area considered in this analysis. The zones were categorized based on spatial coverage of rain gauges, rain gauge location, and elevation difference.

The area includes five different geological groups, which are Baxa Group, Daling Shumar Group, and the Greater, Lesser, and Sub-Himalayan zones. The geology varies from the Baxa Group in the south to the Greater Himalayan Zone in the north. The rocks found in the region are predominantly dark grey to green fine-grained phyllites, slate varying from dark brown to black, and fine-grained, medium-

to thick-bedded quartzite with thin to very thin grey-black fine-grained phyllite interbeds [34,35]. The average thickness of the quartzite is about 100 m, but the individual bands of quartzite range from 10 cm to 2 m. The orientation of the latter is 48° NW, with an average dip 40° towards the slope direction of the slide. The quartzite in the landslide area undergoes brittle deformation with many irregular joints [36]. Figure 2 depicts examples of the damage caused by landslides along the highway.

(**a**)

(**b**)

Figure 2. Landslide damage: (**a**) landslide at 68.1 km along the Samdrup Jongkar–Trashigang (S-T) road (N 26.903°, E 91.505°) (5 July 2016); (**b**) landslide at 93.8 km along the S-T road (N 27.112°, E 91.544°) (17 July 2016) [36].

3. Data

Data from a total of 347 landslides which occurred from 2014 to 2018 were collected from Border Road Organization, Government of India under Project DANTAK (Figure 1b). The data included the dates and geographic coordinates of landslide location based on field observations, interactions with locals, and media reports. The types of movement in the region based on Cruden and Varnes' [37] classification are: debris slides, rock falls, earth flows, and rotational landslides. The field visit conducted in November 2017 revealed the landslides to be shallow, with depths ranging up to few meters, and able to be mapped as single points. The yearly distribution (Figure 3a) of landslides shows that the majority of the landslides occurred in 2017 and 2018, whereas the monthly distribution (Figure 3b) shows that 89% of landslides occurred between the months of June and September. It is

often difficult to determine the rainfall conditions responsible for failures, due to lack of rain gauge density and high distances between rain gauges and landslide points [26,38]. As multiple landslides can occur during a rainfall event, subsequent landslides for a single rainfall event after the initial failure were not considered for the threshold analysis. Thus, in this study, we defined a landslide event as "single landslide-triggering event", in which the landslide events after the initial failure were not considered for threshold estimation [16,26]. A rainfall event was defined as a period of continuous rainfall separated by dry (without rainfall) period. Using all these criteria, the total numbers of rainfall and landslide events during the threshold determination time frame (2014–2017) were 477 and 104, respectively. The zones were divided based on spatial coverage of rain gauges [4,26]. A buffer radius of 15 km around each rain gauge was selected to divide the region into zones. In terms of landslide events for respective zones, half of the landslide events occurred in Zone 2 (49) followed by Zone 3 (40) and Zone 1 (15).

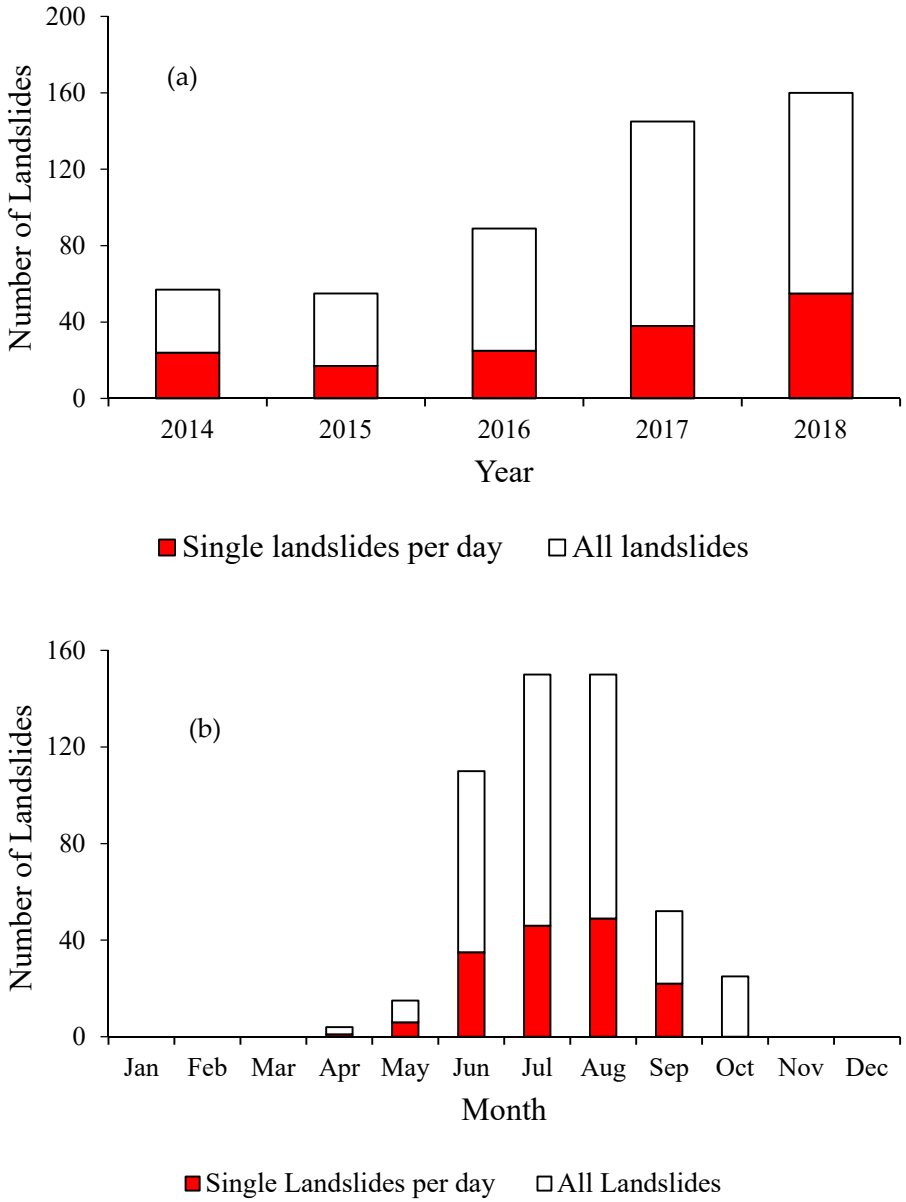


Figure 3. Number of landslides (**a**) per year and (**b**) per month between 2014 and 2018.

The daily rainfall data used for this study (Figure 4) was collected from two rain gauges at Deothang and Kanglung, managed by the National Centre for Hydrology and Meteorology, Bhutan

(http://www.hydromet.gov.bt). The average cumulative yearly rainfalls for Deothang and Kanglung for 2014–2018 were 3495 mm and 1020 mm, respectively, of which 89% occurred during the monsoon season (June–September). The higher precipitation observed in Deothang is due to its location on the windward side of the mountain.

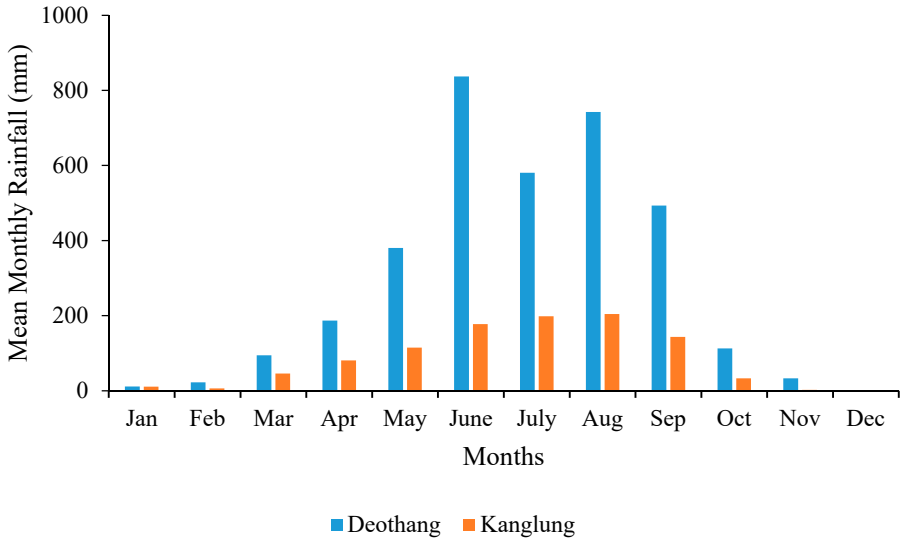


Figure 4. Average monthly rainfall in the study region for both rain gauges (2014–2018).

4. Landslide Temporal Probability Assessment

4.1. Rainfall Threshold Estimation

A rainfall threshold determines the minimum rainfall conditions necessary for landslide initiation in a specific region [39], and various researchers have attempted to quantify thresholds using several approaches [39–41]. A recent review article [40] on rainfall threshold estimation explained in detail the various approaches currently in use, along with their merits and demerits. Of the various techniques, empirically based approaches are widely used because of the simplicity and ease with which they can provide an accurate approximation of minimum precipitation conditions. Various rainfall thresholds using different parameters have been developed, such as ID (intensity–duration) [21], ED (cumulative event rainfall duration) [4], and AD (antecedent rainfall duration) [16,42]. The most commonly used rainfall variables for threshold estimation are daily rainfall, antecedent rainfall, and cumulative rainfall [40]. However, the choice of rainfall variable with which to determine thresholds is primarily dependent on the type of landslides in the region [20].

For the S-T highway, monsoonal rainfall occurs with interruptions and can be characterized mostly low-intensity and long-duration events along with occasional extreme events, making the choice of antecedent rainfall appropriate [30]. Antecedent rainfall is a significant factor for landslide triggering, especially in less impermeable soils, as it lowers soil suction and increases pore water pressure [13]. The use of antecedent rainfall was based on analysis of historical landslide pattern, the field visit, and previous studies conducted in other regions of Bhutan. Although estimation of the number of days to be considered to analyze the effect of antecedent rainfall was a challenge, it has been widely accepted that antecedent rainfall over 15 days to 30 days plays a crucial role for landslide initiation in the Himalayas [43]. The calculation of the antecedent period prior to landslide incidence is usually based on a trial-and-error approach, ranging from 3 days to 120 days [30,44,45].

For this study, the correlation between daily and antecedent rainfall conditions was analyzed for six different time periods (3, 5, 7, 10, 15, and 30 days) (Figure 5a–f). The analysis of the various antecedent rainfall time periods was conducted using the method proposed by Zezere et al. [42]. Blue dots represent the daily rainfall, whereas the orange points depict the antecedent rainfall values for respective time periods. The best discrimination between daily and antecedent conditions, according to

the method suggested by Reference [42], was observed for 30 day antecedent rainfall and was accepted as the metric for threshold calculation.

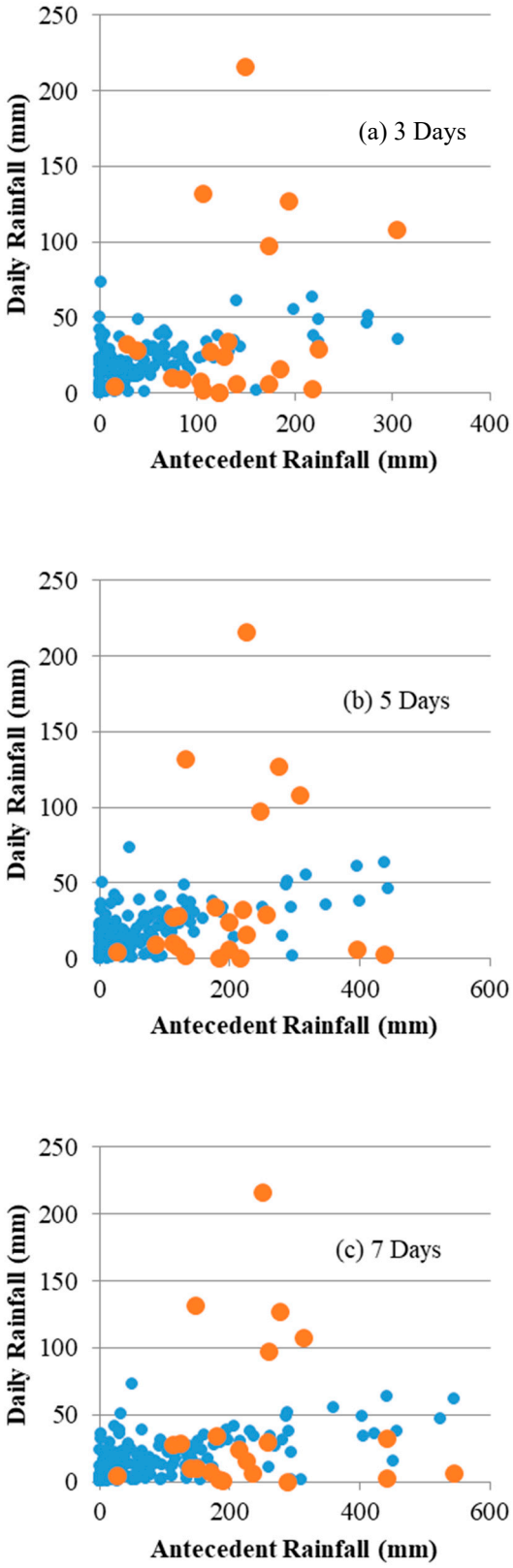


Figure 5. *Cont.*

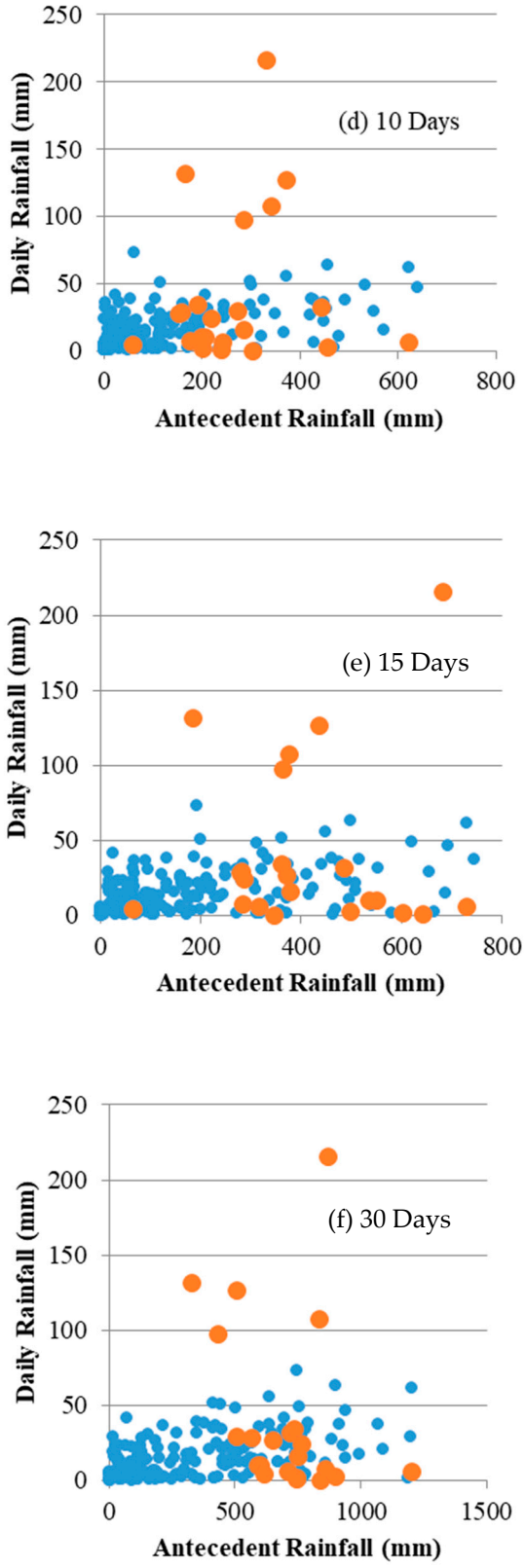


Figure 5. Relationship between daily (blue) and antecedent (orange) rainfall in 2014–2017.

The threshold determination was performed using a scatter chart for daily (R_{TH}) and 30 day antecedent rainfall (R_{30ad}), and was calculated for all three zones. The graph was generated using the rainfall and landslide data from 2014–2017. The threshold equation was obtained by using the lower

end points in the scattered graph [13,16]. The threshold equations of various zones are depicted in Figure 6.

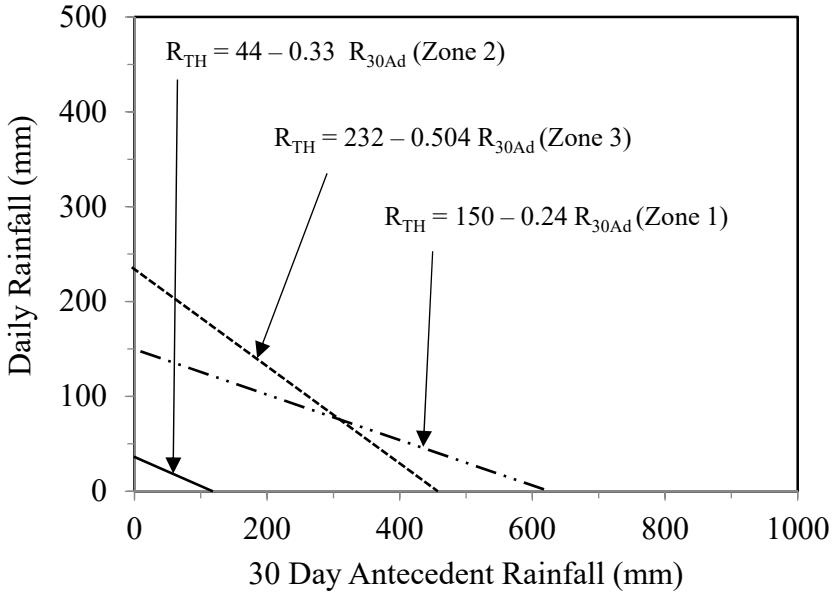


Figure 6. Threshold equation between daily rainfall and 30 day antecedent rainfall for all the three zones.

4.2. Validation of Rainfall Threshold

The significance of any landslide study is determined by the validation of the results obtained. One review of rainfall threshold studies [40] emphasized the importance of validation of rainfall thresholds for conducting landslide studies. The rainfall threshold validation was performed using rainfall and landslide data from 1 January to 31 December 2018. During this period, a total of 52 landslide events occurred, out of which 80% of landslides (41) happened during the monsoon season. The threshold equation for Zone 1 was $R_{TH} = 150 - 0.24R_{30ad}$, and its validation for the monsoon of 2018 is depicted in Figure 7. The threshold exceedance axis depicts the value of R_{TH} with respect to R_{30AD} value for each day, wherein the positive values indicates landslide occurrence. This figure shows that a heavy measure of rainfall occurred, exceeding the threshold. The abrupt increase in the magnitude of daily rainfall or constant rise in 30 day antecedent rainfall is shown by the rise in the threshold curve.

During the validation period, the threshold was exceeded nine times, out of which seven times landslides occurred. No landslides were reported on 2 June and 7 June, even though the threshold was exceeded. This observation can be attributed to the fact that a landslide event does not generally happen with the increase in threshold curve, and sometimes happens a couple of days later as the result of a difference in pore pressure because of changes in the measure of antecedent rainfall. From 1 October to 30 December, there was no threshold exceedance, and no landslides occurred during that period. Similar validations were carried out for Zone 2 and Zone 3 threshold equations using 2018 rainfall and landslide data. These results indicate that the threshold model performed well for landslide forecasting in 2018.

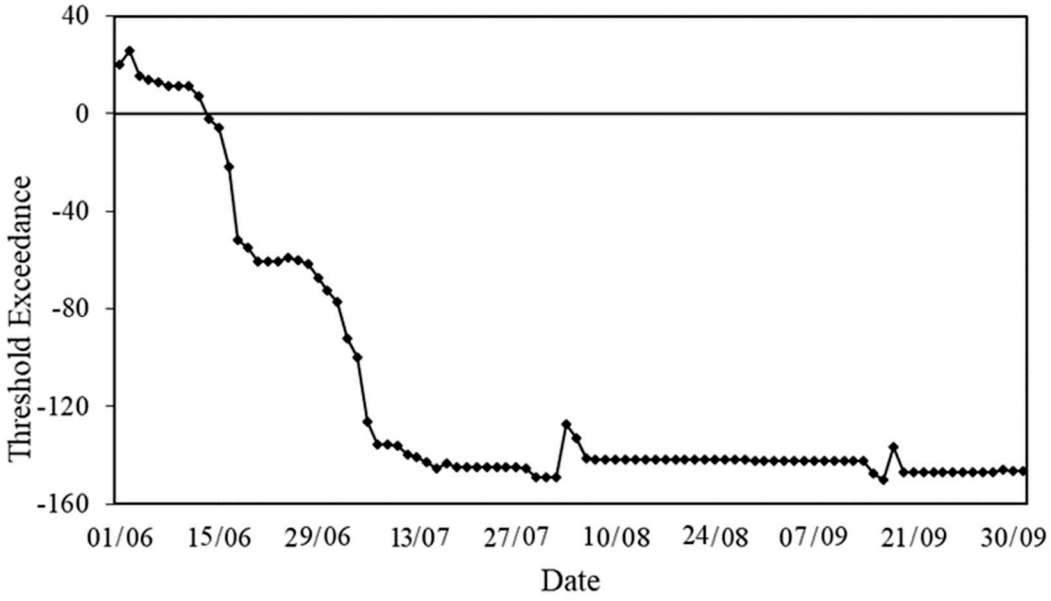


Figure 7. Validation of the threshold equation for Zone 1. The positive values on the Y axis represent exceedance and negative values denote non-exceedance of the threshold.

4.3. Temporal Probability of Landslide Initiation

The temporal probability of landslide incident is determined as the product of annual exceedance probability (AEP) and probability of landslide occurrence [16]. AEP is defined as the probability of the threshold being exceeded in a given year [46], and is determined using the Poisson probability model defined as [16]

$$P(N(t) = n) = e^{-\lambda t}\frac{(-\lambda t)^n}{n!} \quad (1)$$

where N(t) represents the number of landslide incidences during time t and λ is landslide occurrence rate. The exceedance probability for time t is calculated as [13]

$$P[N(t) \geq 1] = 1 - \exp(-t/\mu) \quad (2)$$

where μ is the mean recurrence interval between subsequent landslides determined from landslide inventory.

The determination of temporal probability was based on the following assumptions: (i) probability of landslide incidence is correlated to the probability of rainfall threshold being reached or exceeded, and (ii) landslides will not or will seldom occur when precipitation value is less than the threshold value [13,16,47].

The annual temporal probabilities for different zones of the study region are depicted in Figure 8, and their distribution along with the threshold equations is presented in Table 1. For Zone 1, the threshold value was exceeded 55 times during the simulation period, and out of these 55 cases, landslides were triggered in 16. The estimated probability P[L|(R > RT)] for Zone 1 was 0.29. Similarly, for Zone 2 and 3, the threshold value was exceeded 76 and 64 times in the 4 year period, leading to 31 and 21 landslides being triggered and contributing temporal probability values of 0.41 and 0.33, respectively. The probability of having one or more rainfall events in any given year varied from 0.29 to 0.41. The highest probability values were obtained for Zone 2, followed by Zone 3 and Zone 1. This variation was also observed in the number of landslide incidences for each zone. These precipitation events were also capable of triggering multiple landslides during the monsoon period.

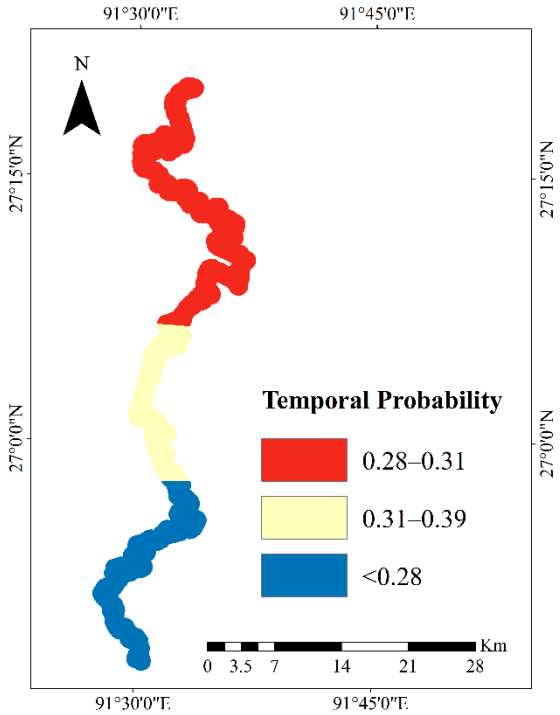


Figure 8. Annual exceedance probability (AEP) of landslide occurrences along the Samdrup Jongkhar–Trashigang Highway.

Table 1. Temporal Probability of Landslide Occurrence.

Area	Threshold Equation (R_T)	Number of Threshold Exceedances	$P(R > R_T)$	Landslide Frequency	$P[L\|(R > R_T)]$	Temporal Probability
Zone 1	$150 - 0.24R_{30ad}$	55	0.99	16	0.29	0.28
Zone 2	$232 - 0.50R_{30ad}$	76	0.99	31	0.40	0.39
Zone 3	$44 - 0.30R_{30ad}$	64	0.99	21	0.32	0.31

5. Landslide Susceptibility Mapping

Landslide susceptibility can be defined as the probability of spatial occurrences of slope failures for a given set of geo-environmental conditions [48], and its determination is one of the crucial steps needed to understand identify potentially landslide-prone sections for any study region. Several studies around the world have been conducted towards the development of landslide susceptibility maps (LSM) using various methods [8]; however, there seems to be no consensus as to the best method for analysis [49]. Aleotti and Chowdhury [50] categorized LSM methods as either quantitative or qualitative. Qualitative models are mostly based on expert opinion, whereas quantitative models are data-driven, which makes them more reliable. The quantitative approaches include several kinds of techniques such, including statistical, deterministic, and other approaches [51–53]. In the case of statistical approaches, it is assumed that the parameters affecting landslide events in the past will be the same in future [54], and these analyses can be categorized into bivariate and multivariate [49]. In bivariate analysis, the factors affecting landslides are compared with landslide inventory data by providing weights based on landslide causative factors. The most frequently used methods in bivariate models are overlay, index-based, and weight-of-evidence analyses [8,51,55]. Bui et al. [6] performed a comparison between a bivariate approach (statistical index) and a multivariate approach (logistic regression) for Vietnam, and found equal forecasting capability. However, one of the main issues with the use of a quantitative approach is the assignment of weights to the landslide-affecting factors [56–58]. The use of GIS has been proven to be a powerful tool with which to validate the significance of factors, and it has been used for multi-criterion decision analysis [49,59]. The decision

analysis technique combines primary- and secondary-level weights for every causative factor, where primary weights are similar to the bivariate approach and secondary weights are expert-opinion-based. For secondary weights, the analytic hierarchy process (AHP) has become popular and has been successfully applied for decision-making systems [60,61]. AHP uses a pairwise relative comparison between every landslide-causative factor. Generally, AHP consists of five key steps: (a) simplify the decision process into its component factors, (b) distribute the factors in a hierarchy process, (c) allocate numerical values to analyze the relative significance of each factor, (d) compose a comparison matrix, and (e) provide weights to every factor by calculating normalized principal eigenvectors [62].

To determine susceptibility, a variety of factors responsible for landslides in the study region were considered. Parameter selection depends on various factors, such as landslide type, data availability and reliability, and adopted methods [63]. For the present study, we used eight landslide-conditioning factors based on the characteristics of the area and prepared from various data sources (Table 2). Figure 9a–f represents all the maps used for the analysis, derived from the Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) with 30 × 30 m resolution, which was the only terrain data source available for this region. The factors with continuous values were reclassified into categories based on Jenks' natural breaks optimization method [64] and developed using ArcGIS 10.4.1.

Table 2. Parameters Used for Landslide Susceptibility Mapping.

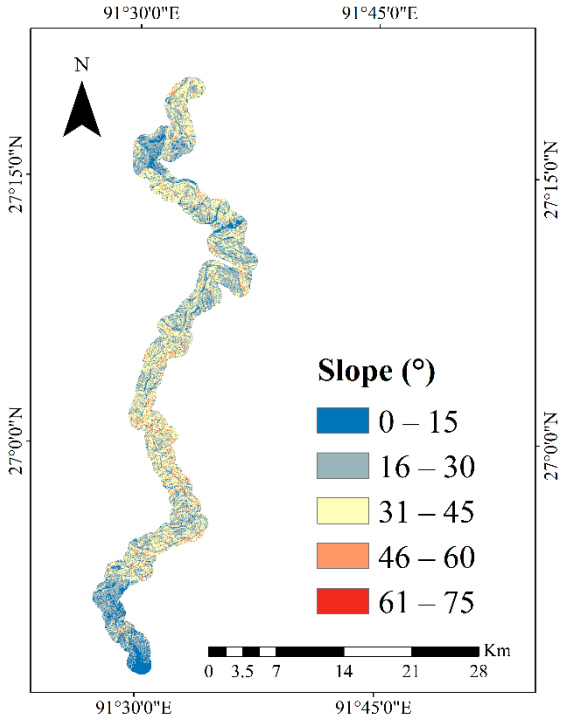
Parameters	Data Source	Explanation	Scale
Slope (°)	SRTM	Derived from raster DEM	1:30,000
Average daily rainfall (mm)	Project DANTAK, Border Road Organization, Govt. of India		1:30,000
Proximity to road (m)	Topographical map	Shape file	1:30,000
Proximity to stream (m)	SRTM	Derived from Raster DEM using the [65] order greater than 5 in vector format	1:30,000
Aspect (N/E/S/W)	SRTM	Derived from raster DEM	1:30,000
Elevation (m)	SRTM	Derived from raster DEM	1:30,000
Landcover	Ministry of Agriculture and Forests, Royal Govt. of Bhutan	Vector data	1:30,000
Geology	Department of Geology and Mines, Royal Govt. of Bhutan	Geological map	1:30,000

The above-mentioned thematic layers were combined by using a weight-of-factors approach determined by AHP to develop the landslide susceptibility map. The use of AHP to develop landslide susceptibility maps has been successfully applied in various regions [61,66,67]. The weights required to carry out AHP were calculated by performing pairwise comparisons for each landslide factor and assigning values from 1 to 9 [63,68–70]. Table 3 shows the pairwise comparison and priority calculation, along with rankings of all indicators. These values were based on an expert's opinion and were placed in n × n matrix, where n is the number of factors.

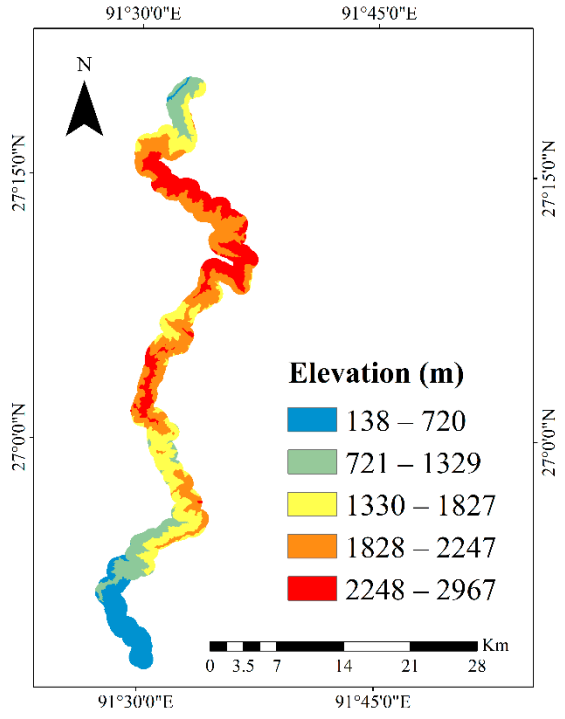
Table 3. Parameter Wise Weights, Matrix, and Consistency Ratio as Determined Using AHP.

Parameters	Slope	Average Daily Rainfall	Proximity to Road	Proximity to Stream	Aspect	Elevation	Landcover	Geology	Weights (%)
Slope	1	0.33	4	2	2	2	3	3	19.2
Average daily rainfall	3	1	4	2	2	3	3	3	26.7
Proximity to road	0.25	0.25	1	0.33	0.33	0.5	0.5	0.33	4.1
Proximity to stream	0.5	0.5	3	1	0.5	1	2	2	11.2
Aspect	0.5	0.5	3	2	1	3	2	2	15.2
Elevation	0.5	0.33	2	1	0.33	1	0.5	0.5	7
Landcover	0.33	0.33	2	0.5	0.5	2	1	1	8
Geology	0.33	0.33	3	0.5	0.5	2	1	1	8.5

Consistency Ratio = 0.039.

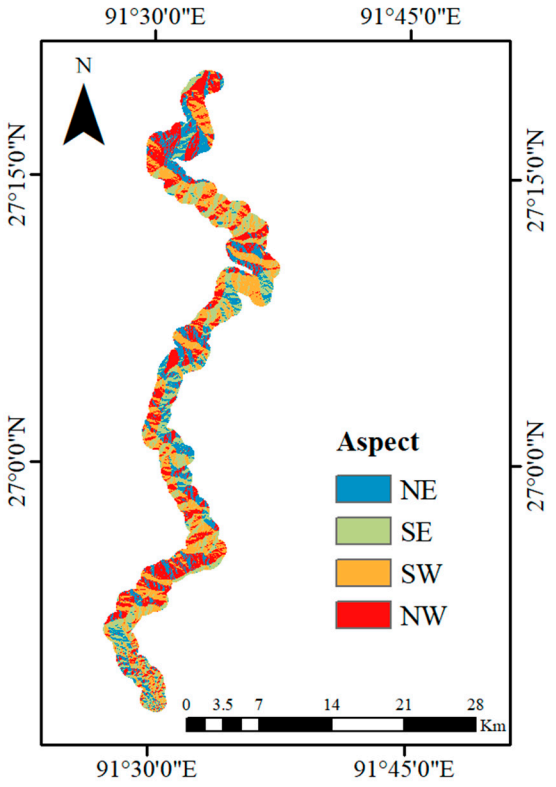


(**a**)

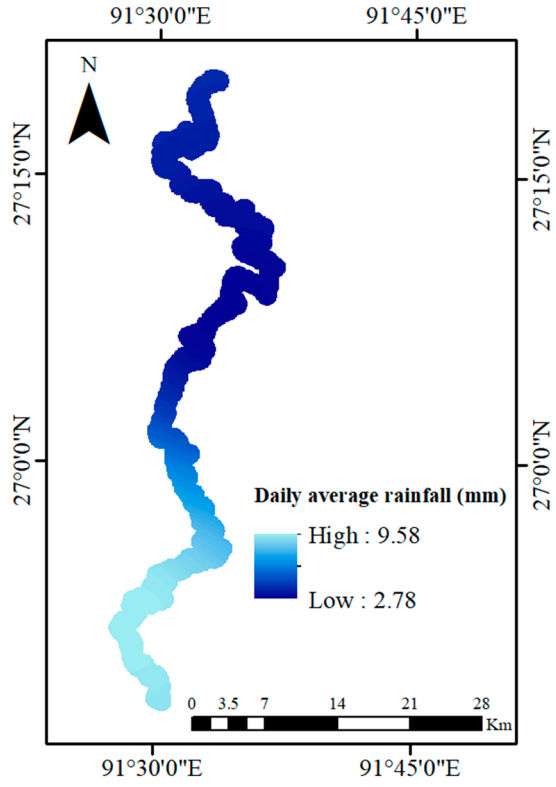


(**b**)

Figure 9. *Cont.*

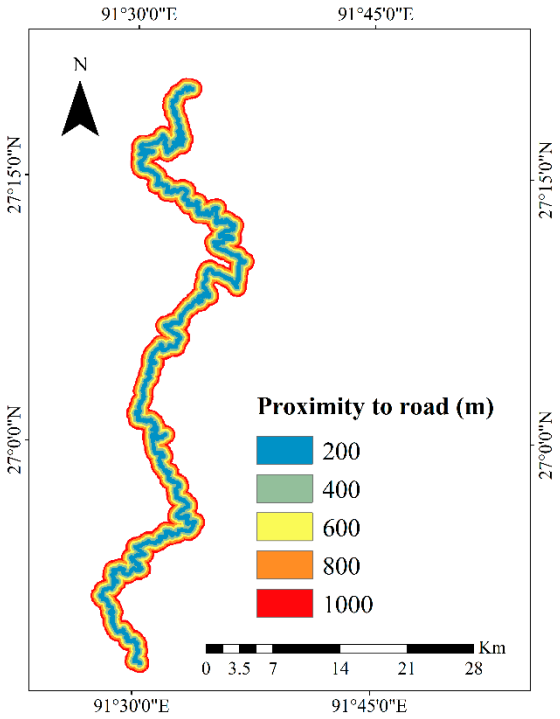


(**c**)

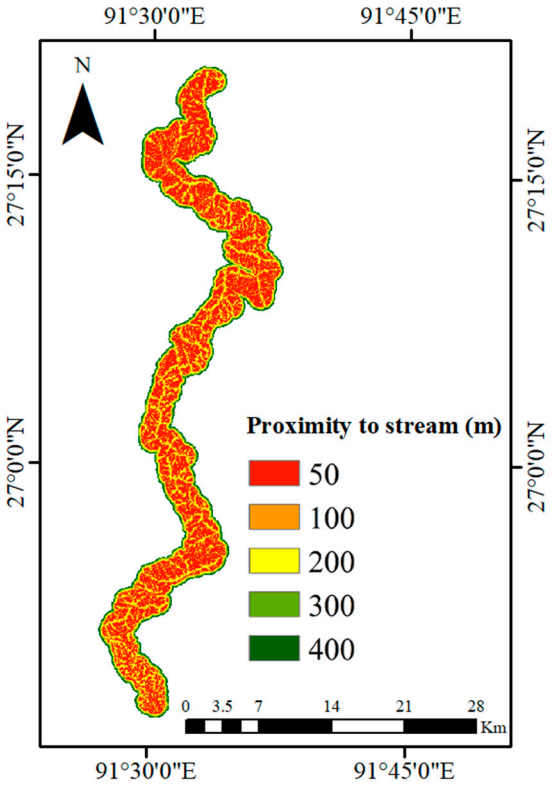


(**d**)

Figure 9. *Cont.*

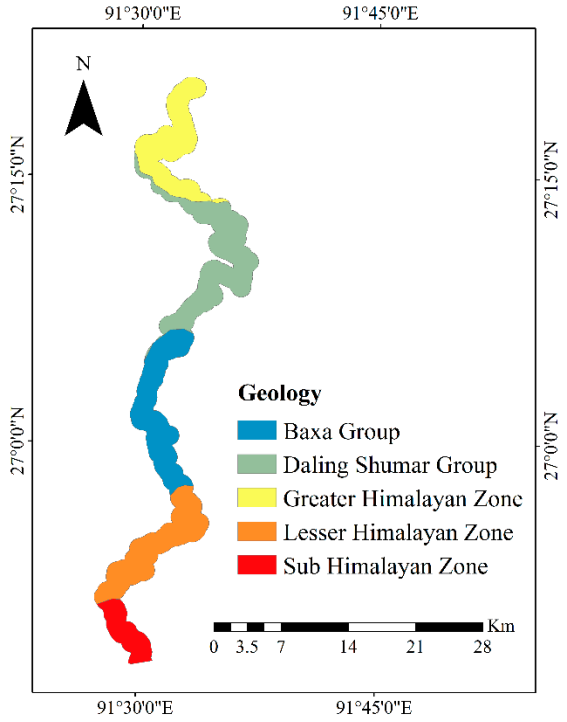


(**e**)

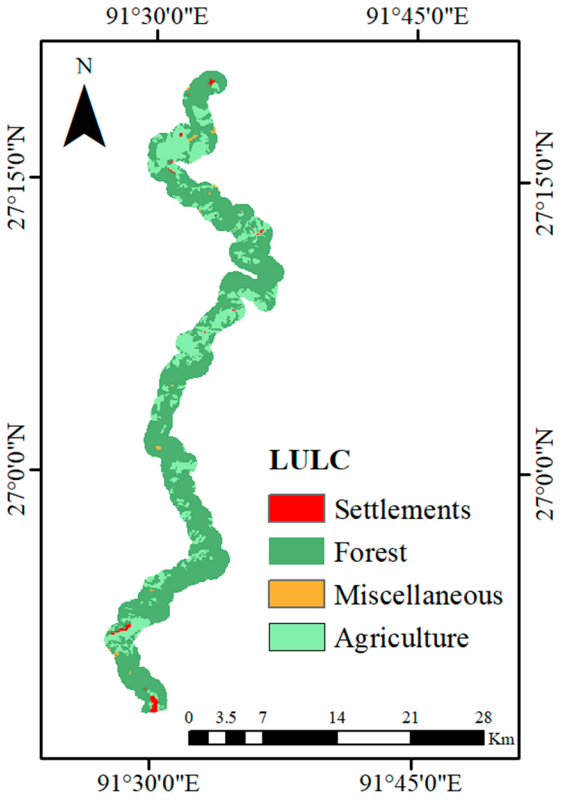


(**f**)

Figure 9. *Cont.*



(**g**)



(**h**)

Figure 9. Landslide-conditioning factor maps: (**a**) slope, (**b**) elevation, (**c**) aspect, (**d**) mean daily rainfall, (**e**) proximity to road, (**f**) proximity to stream, (**g**) geology, and (**h**) land use and land cover (LULC).

The AHP reduced the inconsistencies formed due to the subjectivity of different experts' opinions by computing a consistency index (CI) and consistency ratio (CR), which were determined by

$$CI = (\lambda_{max} - n)/(n - 1) \quad (3)$$

$$CR = CI/RI \quad (4)$$

where λ_{max} represents the largest Eigenvector of the matrix and n represents the total causative factors (order of the matrix) used in the generation of the LSM. RI (random index) is the average value of CI for a randomly generated pairwise matrix and can be accepted only when CR values are less than 10% [71]. For the present study, the average consistency index was estimated for a sample size of 500 and its value was 0.039 (3.9%), which was considered acceptable. Several authors have calculated and estimated different RIs based on various simulation methods and the total number of matrices involved in the process (Table 4). However, we have used Satty's [71] RI values of n = 11 and up to 500 matrices, where the values are 0, 0, 0.58, 0.9, 1.12, 1.24, 1.32, 1.41, 1.45, 1.49, and 1.51.

Table 4. RI Values Obtained by Various Authors (Adopted from Reference [72]).

	[71]	[71]	[73]	[74]	[75]	[76]	[77]	[78]	[79]
	100	500	1000	2500		500		100,000	100,000
3	0.382	0.58	0.5799	0.52	0.5233	0.49	0.500	0.525	0.5245
4	0.946	0.90	0.8921	0.87	0.8860	0.82	0.834	0.882	0.8815
5	1.220	1.12	1.1159	1.10	1.1098	1.03	1.046	1.115	1.1086
6	1.032	1.24	1.2358	1.25	1.2539	1.16	1.178	1.252	1.2479
7	1.468	1.32	1.3322	1.34	1.3451	1.25	1.267	1.341	1.3417
8	1.402	1.41	1.3952	1.40		1.31	1.326	1.404	1.4056
9	1.350	1.45	1.4537	1.45		1.36	1.369	1.452	1.4499
10	1.464	1.49	1.4882	1.49		1.39	1.406	1.484	1.4854
11	1.576	1.51	1.5117			1.42	1.433	1.513	1.5141
12	1.476		1.5356	1.54		1.44	1.456	1.535	1.5365
13	1.564		1.5571			1.46	1.474	1.555	1.5551
14	1.568		1.5714	1.57		1.48	1.491	1.570	1.5713
15	1.586		1.5831			1.49	1.501	1.583	1.5838

The values of CR cannot be negative and can attain a maximum value of 0.3. Values of CR less than 0.1 are considered acceptable; if this is not achieved, new attempts are made until the value is acceptable [80]. However, the values of CR are dependent on the analysis type and the number of criteria being considered. In some cases, CR > 0.1 may not be considered critical, and values of CR ranging from 0.15 to 0.3 can also be considered acceptable. A matrix will be considered consistent, according to Saaty [71], if

$$\lambda max < n + 0.1((\lambda max) - n) \quad (5)$$

Finally, all the landslide-causative factors and classes were integrated by a method of weighted overlay in ArcGIS to generate the landslide susceptibility index (LSI).

$$LSI = \sum_{k=1}^{n} W_k W_{jk} \quad (6)$$

where W_k is the weight of the causative factor, W_{jk} is the rank value for factor class j of causative factor k, and n represents the total causative factors selected. Ranks of criteria were calculated based on priority or weight values. The highest priority considered was Rank 1, while the lowest priority considered was the last rank in the AHP.

Model Development and Validation

The landslide inventory data was randomly categorized into two datasets, i.e., training (70%) and testing (30%). The landslide susceptibility map based on 8 causative factors using AHP is depicted in Figure 10. The map was divided into 4 classes: very low (0–0.24), low (0.25–0.49), moderate (0.5–0.74) and high (0.75–1) according to natural breaks to define the class intervals in the susceptibility map. From total of 242 landslides, more than 78.1% falls under the high zone. Whereas 11.98% falls under the moderate zone and combined 9.92% falls under the low and very low zones (Table 5).

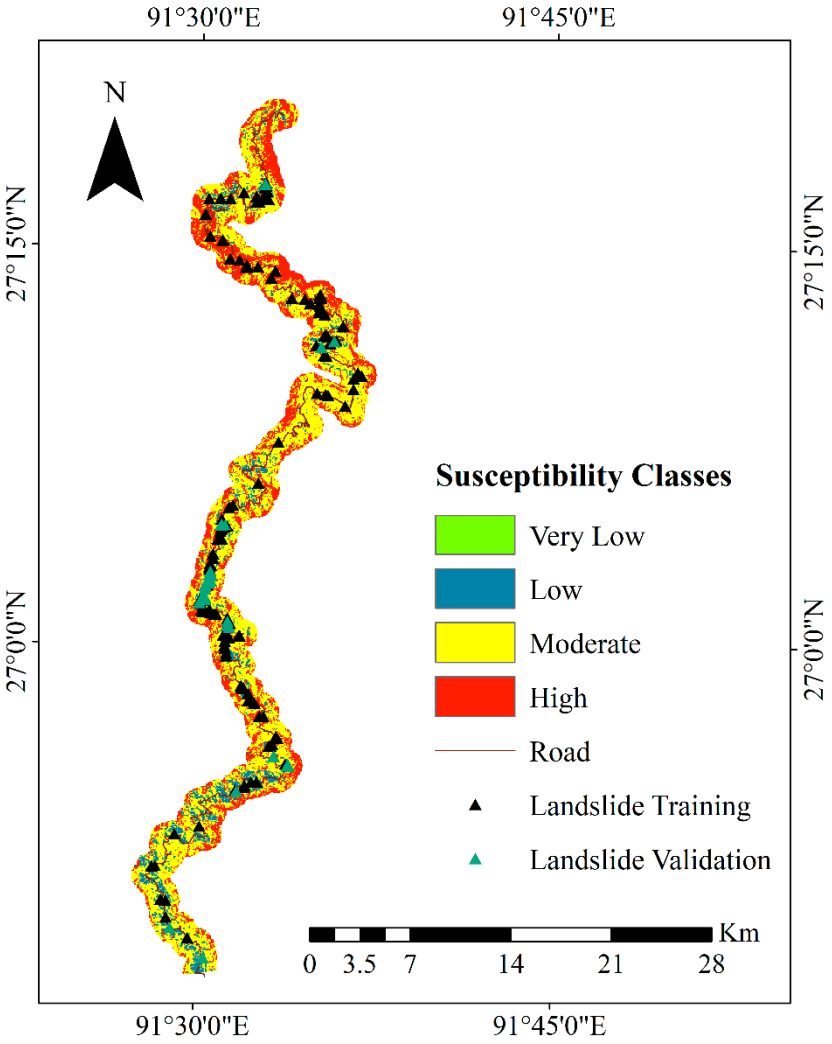


Figure 10. Landslide Susceptibility Map of the Study Region.

Table 5. Landslide Susceptibility Results for the Samdrup Jongkhar–Trashigang Highway Region.

Inventory		Susceptibility							
Zone	Landslides	High		Moderate		Low		Very Low	
		(%)	(km^2)	(%)	(km^2)	(%)	(km^2)	(%)	(km^2)
Zone 1	39	21.23	13.84	59.90	39.05	18.71	12.19	0.15	0.10
Zone 2	119	36.6	38.82	55.7	59.08	6.8	7.25	0.9	0.94
Zone 3	84	28.31	13.69	63.45	30.67	8.06	3.89	0.16	0.08

Validation of the susceptibility maps was performed for randomly selected data from the inventory data using receiver operating characteristics (ROC). This is an effective way to analyze the quality of

predictive techniques [81]. The ROC curve is plotted between true positive rate (sensitivity) on the Y axis against false positive rate (specificity) on the X axis. The terms "sensitivity" and "specificity", which are used to plot ROC curves, are defined as follows.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \tag{8}$$

where true positive (TP) is the number of actual landslides predicted correctly, and true negative (TN) is the total number of non-occurring landslides predicted correctly. False positive (FP) is the number of actual landslides inaccurately predicted as non-occurring landslides, and false negative (FN) is the number of non-occurring landslides inaccurately predicted as actual landslides. [82]. The area under the ROC curve (AUC) was also used to determine the quality of the prediction by analyzing the model's ability to forecast the occurrence or nonoccurrence of predefined events [83]. The results of the success rate curve of the AHP model had an AUC of 0.798, corresponding to a prediction accuracy of 79.8% (Figure 11).

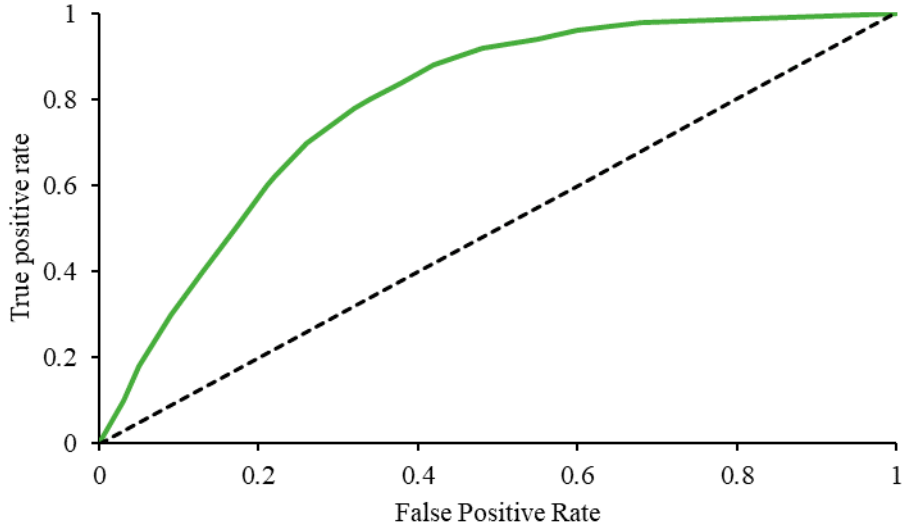


Figure 11. Receiver operating characteristic (ROC) curve of the susceptibility map by AHP.

The results of pairwise comparison, priority estimation, and ranking of all the criteria could be applied to other study areas for susceptibility assessment. Several high-resolution satellite image datasets are required to better understand the locations and perform these assessments. In cases of unavailability of high-resolution satellite images, Google Earth images could be useful for preparation of some indicators. The use of Google Earth images could also be helpful for accurate identification of landslide locations and conduction of extensive field studies. All the requirements for more accurate analysis depend on study location tectonics, data availability, and proper methods. Therefore, based on local geology and tectonic conditions, these results could be transferable and applicable in other locations in both small- and large-scale areas.

6. Conclusions

Landslides are the most frequently occurring natural hazards, especially in the Himalayan regions, which suffer from heavy monsoonal rainfall and subsequent landslides. In this study, the temporal probability of landslide events was determined using rainfall and landslide data from 2014–2017 along Samdrup Jongkhar–Trashigang highway in East Bhutan. The highway was divided into three zones based on land use, topography, and rain gauge coverage for the determination of temporal probability.

Thereafter, a landslide susceptibility map was developed using AHP. The results of temporal probability were validated with landslide event dataset of 2018 to understand the applicability of the model. The conclusions from the study can be summarized as follows.

(1) Threshold determination was performed considering the antecedent rainfall duration approach. The selection of antecedent rainfall was based on previous studies which have highlighted its significance, especially in the Himalayan region and in Bhutan. A trial-and-error approach was adopted to determine the number of antecedent days required for landslide initiation, and a 30 day antecedent rainfall was adopted for threshold determination.
(2) The temporal probability of landslide occurrence was determined based on the Poisson model, and the validation results based on 2018 data revealed that the model performed well.
(3) The landslide susceptibility map of the region was developed using AHP classified into four categories (Very Low, Low, Moderate, and High), and the results showed that 78% of the region falls under into high-susceptibility zone. The performance of the model was assessed using the area under ROC and an accuracy of 79.8% was achieved. However, due to the dynamic nature of land use and rainfall patterns, the susceptibility map will need to be updated from time to time.

The present study on rainfall threshold estimation and the development of a susceptibility map for eastern Bhutan along Samdrup Jongkhar–Trashigang highway is an important study in the context of the Bhutan Himalayas, for which study on both these aspects is lacking. The current study can be regarded as a preliminary step towards risk management, which could be supported by conducting a hazard and vulnerability assessment of the region. The temporal probabilities determined can be integrated with the susceptibility map to obtain a landslide hazard map. However, the results might be improved by increasing the number of landslide events and using precipitation data with a higher temporal resolution. The results from the present study also could prove helpful to civic authorities in identifying key sections of the road which are most vulnerable to landslides, and undertaking strict measures to prevent slope failures, strengthening the transportation network and saving human lives.

Author Contributions: Conceptualization, A.D. and R.S.; methodology, A.D. and R.J.; data curation, D.D.; writing—original draft preparation, A.D. and R.S.; writing—review and editing, B.P. and A.M.A.; supervision, B.P. and R.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS) in the University of Technology Sydney (UTS) under Grants 321740.2232335 and 321740.2232357; Grant 321740.2232424, and Grant 321740.2232452. This research is also supported by Researchers Supporting Project number RSP-2019/14, King Saud University, Riyadh, Saudi Arabia. Also, the research was funded from BRACE project (NERC/GCRF NE/P016219/1).

Acknowledgments: The authors are thankful to National Center for Hydrology and Meteorology, Royal Government of Bhutan for providing rainfall data and the Border Roads Organization (Project DANTAK), Government of India for providing landslide data. Authors are also thankful to staff of College of Science and Technology, Royal University of Bhutan, who helped directly or indirectly while carrying out the present study. We also thank the editor and the reviewers for reviewing and suggesting valuable modifications, which has helped in improving the quality of this paper.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Petley, D. Global patterns of loss of life from landslides. *Geology* **2012**, *40*, 927–930. [CrossRef]
2. Froude, M.J.; Petley, D.N. Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 2161–2181. [CrossRef]
3. Stanley, T.; Kirschbaum, D.B. A heuristic approach to global landslide susceptibility mapping. *Nat. Hazards* **2017**, *87*, 145–164. [CrossRef]
4. Dikshit, A.; Sarkar, R.; Pradhan, B.; Acharya, S.; Dorji, K. Estimating Rainfall Thresholds for Landslide Occurrence in the Bhutan Himalayas. *Water* **2019**, *11*, 1616. [CrossRef]

5. Sarkar, R.; Dorji, K. Determination of the Probabilities of Landslide Events—A Case Study of Bhutan. *Hydrology* **2019**, *6*, 52. [CrossRef]
6. Bui, D.T.; Löfman, O.; Revhaug, I.; Dick, O. Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. *Nat. Hazards* **2011**, *59*, 1413–1444. [CrossRef]
7. Qiu, H.; Cui, Y.; Hu, S.; Yang, D.; Pei, Y.; Yang, W. Temporal and spatial distributions of landslides in the Qinba Mountains, Shaanxi Province, China. *Geomat. Nat. Hazards Risk* **2019**, *10*, 599–621. [CrossRef]
8. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* **2018**, *180*, 60–91. [CrossRef]
9. Althuwaynee, O.F.; Pradhan, B.; Lee, S. Application of an evidential belief function model in landslide susceptibility mapping. *Comput. Geosci.* **2012**, *44*, 120–135. [CrossRef]
10. Pradhan, B.; Lee, S. Landslide susceptibility assessment and factor effect analysis: Backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ. Model. Softw.* **2010**, *25*, 747–759. [CrossRef]
11. Pradhan, B.; Lee, S.; Buchroithner, M.F. A GIS-based back-propagation neural network model and its cross-application and validation for landslide susceptibility analyses. *Comput. Environ. Urban Syst.* **2010**, *34*, 216–235. [CrossRef]
12. Jaiswal, P.; Van Westen, C.J.; Jetten, V. Quantitative landslide hazard assessment along a transportation corridor in southern India. *Eng. Geol.* **2010**, *116*, 236–250. [CrossRef]
13. Bui, D.T.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, Ø.B. Regional prediction of landslide hazard using probability analysis of intense rainfall in the Hoa Binh province, Vietnam. *Nat. Hazards* **2013**, *66*, 707–730.
14. Afungang, R.N.; Bateira, C.V. Temporal probability analysis of landslides triggered by intense rainfall in the Bamenda Mountain Region, Cameroon. *Environ. Earth Sci.* **2016**, *75*, 1032. [CrossRef]
15. Saez, J.L.; Corona, C.; Stoffel, M.; Schoeneich, P.; Berger, F. Probability maps of landslide reactivation derived from tree-ring records: Pra Bellon landslide, southern French Alps. *Geomorphology* **2012**, *138*, 189–202. [CrossRef]
16. Jaiswal, P.; Van Westen, C.J. Estimating temporal probability for landslide initiation along transportation routes based on rainfall thresholds. *Geomorphology* **2009**, *112*, 96–105. [CrossRef]
17. Corominas, J.; Moya, J. A review of assessing landslide frequency for hazard zoning purposes. *Eng. Geol.* **2008**, *102*, 193–213. [CrossRef]
18. Das, I.; Stein, A.; Kerle, N.; Dadhwal, V.K. Probabilistic landslide hazard assessment using homogeneous susceptible units (HSU) along a national highway corridor in the northern Himalayas, India. *Landslides* **2011**, *8*, 293–308. [CrossRef]
19. Segoni, S.; Lagomarsino, D.; Fanti, R.; Moretti, S.; Casagli, N. Integration of rainfall thresholds and susceptibility maps in the Emilia Romagna (Italy) regional-scale landslide warning system. *Landslides* **2015**, *12*, 773–785. [CrossRef]
20. Martelloni, G.; Segoni, S.; Fanti, R.; Catani, F. Rainfall thresholds for the forecasting of landslide occurrence at regional scale. *Landslides* **2012**, *9*, 485–495. [CrossRef]
21. Giannecchini, R.; Galanti, Y.; Avanzi, G.D. Critical rainfall thresholds for triggering shallow landslides in the Serchio River Valley (Tuscany, Italy). *Nat. Hazards Earth Syst. Sci.* **2012**, *12*, 829–842. [CrossRef]
22. Schmidt, M.; Glade, T. Linking global circulation model outputs to regional geomorphic models: A case study of landslide activity in New Zealand. *Clim. Res.* **2003**, *25*, 135–150. [CrossRef]
23. Althuwaynee, O.F.; Pradhan, B.; Ahmad, N. Estimation of rainfall threshold and its use in landslide hazard mapping of Kuala Lumpur metropolitan and surrounding areas. *Landslides* **2015**, *12*, 861–875. [CrossRef]
24. Dikshit, A.; Satyam, D.N. Estimation of rainfall thresholds for landslide occurrences in Kalimpong, India. *Innov. Infrastruct. Solut.* **2018**, *3*, 24. [CrossRef]
25. Gabet, E.J.; Burbank, D.W.; Putkonen, J.K.; Pratt-Sitaula, B.A.; Ojha, T. Rainfall thresholds for landsliding in the Himalayas of Nepal. *Geomorphology* **2004**, *63*, 131–143. [CrossRef]
26. Gariano, S.L.; Sarkar, R.; Dikshit, A.; Dorji, K.; Brunetti, M.T.; Peruccacci, S.; Melillo, M. Automatic calculation of rainfall thresholds for landslide occurrence in Chukha Dzongkhag, Bhutan. *Bull. Eng. Geol. Environ.* **2019**, *78*, 4325–4332. [CrossRef]
27. Crosta, G.B.; Frattini, P. Distributed modelling of shallow landslides triggered by intense rainfall. *Nat. Hazards Earth Syst. Sci.* **2003**, *3*, 81–93. [CrossRef]

28. Jemec, M.; Komac, M. Rainfall patterns for shallow landsliding in perialpine Slovenia. *Nat. Hazards* **2013**, *67*, 1011–1023. [CrossRef]
29. Bui, D.T.; Pradhan, B.; Löfman, O.; Revhaug, I.; Dick, O.B. Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): A comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. *Catena* **2012**, *96*, 28–40.
30. Kanungo, D.; Sharma, S. Rainfall thresholds for prediction of shallow landslides around Chamoli-Joshimath region, Garhwal Himalayas, India. *Landslides* **2014**, *11*, 629–638. [CrossRef]
31. Frattini, P.; Crosta, G.; Sosio, R. Approaches for defining thresholds and return periods for rainfall-triggered shallow landslides. *Hydrol. Process.* **2009**, *23*, 1444–1460. [CrossRef]
32. Sarkar, R.; Dikshit, A.; Hazarika, H.; Yamada, K.; Subba, K. Probabilistic Rainfall Thresholds for Landslide Occurrences in Bhutan. *Int. J. Recent Technol. Eng.* **2019**, *8*. [CrossRef]
33. Sameen, M.I.; Sarkar, R.; Pradhan, B.; Drukpa, D.; Alamri, A.M.; Park, H.-J. Landslide spatial modelling using unsupervised factor optimisation and regularised greedy forests. *Comput. Geosci.* **2019**. [CrossRef]
34. Long, S.; McQuarrie, N.; Tobgay, T.; Grujic, D.; Hollister, L. Geologic Map of Bhutan. *J. Maps* **2011**, *7*, 184–192. [CrossRef]
35. McQuarrie, N.; Long, S.; Tobgay, T.; Nesbit, J.; Gehrels, G.; Ducea, M. Documenting basin scale, geometry and provenance through detrital geochemical data: Lessons from the Neoproterozoic to Ordovician Lesser, Greater, and Tethyan Himalayan strata of Bhutan. *Gondwana Res.* **2013**, *23*, 1491–1510. [CrossRef]
36. Phuntso, Y.; Wangda, U.; Tenzin, T. *Monitoring of Landslide at Arong, Moshi and Phongmey Using GPS*; Department of Geology and Mines, Ministry of Economic Affairs, Royal Government of Bhutan: Thimphu, Bhutan, 2018.
37. Cruden, D.M.; Varnes, D.J. Chapter 3—Landslide types and processes. In *Landslides: Investigation and Mitigation*; Transportation Research Board: Washington, DC, USA, 1996.
38. Gariano, S.L.; Brunetti, M.; Iovine, G.; Melillo, M.; Peruccacci, S.; Terranova, O.; Vennari, C.; Guzzetti, F. Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy. *Geomorphology* **2015**, *228*, 653–665. [CrossRef]
39. Guzzetti, F.; Peruccacci, S.; Rossi, M.; Stark, C.P. Rainfall thresholds for the initiation of landslides in central and southern Europe. *Theor. Appl. Clim.* **2007**, *98*, 239–267. [CrossRef]
40. Segoni, S.; Piciullo, L.; Gariano, S.L. A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides* **2018**, *15*, 1483–1501. [CrossRef]
41. Segoni, S.; Rosi, A.; Lagomarsino, D.; Fanti, R.; Casagli, N. Brief communication: Using averaged soil moisture estimates to improve the performances of a regional-scale landslide early warning system. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 807–812. [CrossRef]
42. Zêzere, J.L.; Trigo, R.M.; Trigo, I.F. Shallow and deep landslides induced by rainfall in the Lisbon region (Portugal): Assessment of relationships with the North Atlantic Oscillation. *Nat. Hazards Earth Syst. Sci.* **2005**, *5*, 331–344. [CrossRef]
43. Mathew, J.; Babu, D.G.; Kundu, S.; Kumar, K.V.; Pant, C.C. Integrating intensity–duration-based rainfall threshold and antecedent rainfall-based probability estimate towards generating early warning for rainfall-induced landslides in parts of the Garhwal Himalaya, India. *Landslides* **2014**, *11*, 575–588. [CrossRef]
44. Glade, T.; Crozier, M.; Smith, P. Applying Probability Determination to Refine Landslide-triggering Rainfall Thresholds Using an Empirical "Antecedent Daily Rainfall Model". *Pure Appl. Geophys.* **2000**, *157*, 1059–1079. [CrossRef]
45. Dahal, R.K.; Hasegawa, S. Representative rainfall thresholds for landslides in the Nepal Himalaya. *Geomorphology* **2008**, *100*, 429–443. [CrossRef]
46. Fell, R.; Ho, K.K.; Lacasse, S.; Leroi, E. A framework for landslide risk assessment and management. In *Landslide Risk Management*; CRC Press: Boca Raton, FL, USA, 2005; pp. 13–36.
47. Chleborad, A.F.; Baum, R.L.; Godt, J.W. Rainfall thresholds for forecasting landslides in the Seattle, Washington, area: Exceedance and probability. In *US Geological Survey Open-File Report*; US Geological Survey: Reston, VA, USA, 2006; Volume 1064, p. 31.
48. Guzzetti, F.; Reichenbach, P.; Cardinali, M.; Galli, M.; Ardizzone, F. Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* **2005**, *72*, 272–299. [CrossRef]
49. Goetz, J.; Brenning, A.; Petschko, H.; Leopold, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* **2015**, *81*, 1–11. [CrossRef]

50. Aleotti, P.; Chowdhury, R. Landslide hazard assessment: Summary review and new perspectives. *Bull. Int. Assoc. Eng. Geol.* **1999**, *58*, 21–44. [CrossRef]
51. Pawluszek, K.; Borkowski, A. Impact of DEM-derived factors and analytical hierarchy process on landslide susceptibility mapping in the region of Rożnów Lake, Poland. *Nat. Hazards* **2017**, *86*, 919–952. [CrossRef]
52. Pourghasemi, H.R.; Rahmati, O. Prediction of the landslide susceptibility: Which algorithm, which precision? *Catena* **2018**, *162*, 177–192. [CrossRef]
53. Bui, D.T.; Shahabi, H.; Omidvar, E.; Shirzadi, A.; Geertsema, M.; Clague, J.J.; Khosravi, K.; Pradhan, B.; Pham, B.T.; Chapi, K.; et al. Shallow Landslide Prediction Using a Novel Hybrid Functional Machine Learning Algorithm. *Remote Sens.* **2019**, *11*, 931. [CrossRef]
54. Guzzetti, F.; Carrara, A.; Cardinali, M.; Reichenbach, P. Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* **1999**, *31*, 181–216. [CrossRef]
55. Kavzoglu, T.; Sahin, E.K.; Colkesen, I. Selecting optimal conditioning factors in shallow translational landslide susceptibility mapping using genetic algorithm. *Eng. Geol.* **2015**, *192*, 101–112. [CrossRef]
56. Tien Bui, D.; Tuan, T.A.; Hoang, N.-D.; Thanh, N.Q.; Nguyen, D.B.; Van Liem, N.; Pradhan, B. Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. *Landslides* **2017**, *14*, 447–458. [CrossRef]
57. Lagomarsino, D.; Tofani, V.; Segoni, S.; Catani, F.; Casagli, N. A Tool for Classification and Regression Using Random Forest Methodology: Applications to Landslide Susceptibility Mapping and Soil Thickness Modeling. *Environ. Model. Assess.* **2017**, *22*, 201–214. [CrossRef]
58. Segoni, S.; Tofani, V.; Rosi, A.; Catani, F.; Casagli, N. Combination of Rainfall Thresholds and Susceptibility Maps for Dynamic Landslide Hazard Assessment at Regional Scale. *Front. Earth Sci.* **2018**, *6*, 85. [CrossRef]
59. Ahmed, B. Landslide susceptibility mapping using multi-criteria evaluation techniques in Chittagong Metropolitan Area, Bangladesh. *Landslides* **2015**, *12*, 1077–1095. [CrossRef]
60. El Jazouli, A.; Barakat, A.; Khellouk, R. GIS-multicriteria evaluation using AHP for landslide susceptibility mapping in Oum Er Rbia high basin (Morocco). *Geoenviron. Disasters* **2019**, *6*, 3. [CrossRef]
61. Shahabi, H.; Hashim, M. Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment. *Sci. Rep.* **2015**, *5*, 9899. [CrossRef]
62. Kayastha, P.; Dhital, M.; De Smedt, F. Application of the analytical hierarchy process (AHP) for landslide susceptibility mapping: A case study from the Tinau watershed, west Nepal. *Comput. Geosci.* **2013**, *52*, 398–408. [CrossRef]
63. Hasekioğulları, G.D.; Ercanoglu, M. A new approach to use AHP in landslide susceptibility mapping: A case study at Yenice (Karabuk, NW Turkey). *Nat. Hazards* **2012**, *63*, 1157–1179. [CrossRef]
64. Hong, H.; Naghibi, S.A.; Pourghasemi, H.R.; Pradhan, B. GIS-based landslide spatial modeling in Ganzhou City, China. *Arab. J. Geosci.* **2016**, *9*, 112. [CrossRef]
65. Crovelli, R.A. *Probability Models for Estimation of Number and Costs of Landslides*; US Geological Survey: Reston, VA, USA, 2000.
66. Palchaudhuri, M.; Biswas, S. Application of AHP with GIS in drought risk assessment for Puruliya district, India. *Nat. Hazards* **2016**, *84*, 1905–1920. [CrossRef]
67. Pourghasemi, H.R.; Pradhan, B.; Gokceoglu, C. Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. *Nat. Hazards* **2012**, *63*, 965–996. [CrossRef]
68. Mokarram, M.; Zarei, A.R. Landslide Susceptibility Mapping Using Fuzzy-AHP. *Geotech. Geol. Eng.* **2018**, *36*, 3931–3943. [CrossRef]
69. Saaty, T.L. A scaling method for priorities in hierarchical structures. *J. Math. Psychol.* **1977**, *15*, 234–281. [CrossRef]
70. Saaty, T.L.; Vargas, L.G. Diagnosis with Dependent Symptoms: Bayes Theorem and the Analytic Hierarchy Process. *Oper. Res.* **1998**, *46*, 491–502. [CrossRef]
71. Saaty, T.L. *Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process*; RWS Publications: Pittsburgh, PA, USA, 2000; Volume 6.
72. Alonso, J.A.; Lamata, M.T. Consistency in the analytic hierarchy process: A new approach. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* **2006**, *14*, 445–459. [CrossRef]

73. Golden, B.L.; Wang, Q. An Alternate Measure of Consistency. In *The Analytic Hierarchy Process*; Springer Science and Business Media: Berlin/Heidelberg, Germany, 1989; pp. 68–81.
74. Lane, E.F.; Verdini, W.A. A Consistency Test for AHP Decision Makers. *Decis. Sci.* **1989**, *20*, 575–590. [CrossRef]
75. Forman, E.H. Random indices for incomplete pairwise comparison matrices. *Eur. J. Oper. Res.* **1990**, *48*, 153–155. [CrossRef]
76. Noble, E.E.; Sanchez, P.P. A note on the information content of a consistent pairwise comparison judgment matrix of an AHP decision maker. *Theory Decis.* **1993**, *34*, 99–108. [CrossRef]
77. Tummala, V.R.; Wan, Y.-W. On the mean random inconsistency index of analytic hierarchy process (AHP). *Comput. Ind. Eng.* **1994**, *27*, 401–404. [CrossRef]
78. Aguarón, J.; Moreno-Jiménez, J.M. The geometric consistency index: Approximated thresholds. *Eur. J. Oper. Res.* **2003**, *147*, 137–145. [CrossRef]
79. Alonso, J.; Lamata, M. Estimation of the random index in the analytic hierarchy process. *Proc. Proc. Inf. Process. Manag. Uncertain. Knowl.-Based Syst.* **2014**, *1*, 317–322.
80. Jena, R.; Pradhan, B.; Beydoun, G.; Nizamuddin; Ardiansyah; Sofyan, H.; Affan, M. Integrated model for earthquake risk assessment using neural network and analytic hierarchy process: Aceh province, Indonesia. *Geosci. Front.* **2019**. [CrossRef]
81. Fan, W.; Wei, X.-S.; Cao, Y.-B.; Zheng, B. Landslide susceptibility assessment using the certainty factor and analytic hierarchy process. *J. Mt. Sci.* **2017**, *14*, 906–925. [CrossRef]
82. Sangchini, E.K.; Emami, S.N.; Tahmasebipour, N.; Pourghasemi, H.R.; Naghibi, S.A.; Arami, S.A.; Pradhan, B. Assessment and comparison of combined bivariate and AHP models with logistic regression for landslide susceptibility mapping in the Chaharmahal-e-Bakhtiari Province, Iran. *Arab. J. Geosci.* **2016**, *9*. [CrossRef]
83. Devkota, K.C.; Regmi, A.D.; Pourghasemi, H.R.; Yoshida, K.; Pradhan, B.; Ryu, I.C.; Dhital, M.R.; Althuwaynee, O.F. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. *Nat. Hazards* **2013**, *65*, 135–165. [CrossRef]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Journal of The Electrochemical Society

The Electrochemical Society
Advancing solid state & electrochemical science & technology

OPEN ACCESS

Review—Textile Based Chemical and Physical Sensors for Healthcare Monitoring

To cite this article: Amir Hatamie *et al* 2020 *J. Electrochem. Soc.* **167** 037546

View the article online for updates and enhancements.

This content was downloaded from IP address 14.139.251.100 on 27/01/2020 at 06:15

Journal of The Electrochemical Society, 2020 **167** 037546

Review—Textile Based Chemical and Physical Sensors for Healthcare Monitoring

Amir Hatamie,[1,2,z] Shayan Angizi,[1] Saurabh Kumar,[3,4] Chandra Mouli Pandey,[5] Abdolreza Simchi,[1] Magnus Willander,[2] and Bansi D. Malhotra[4,z]

[1]*Nanostructured & Advanced Materials Laboratory, Department of Materials Science and Engineering, Sharif University of Technology, Tehran 1458889694, Iran*
[2]*Department of Science & Technology, Linköping University, ITN-Campus, SE-60174 Norrköping, Sweden*
[3]*Centre for Nano Science and Engineering (CeNSE), Indian Institute of Science, Bangalore-560012, India*
[4]*Department of Biotechnology, Delhi Technological University, Delhi 110042, India*
[5]*Department of Applied Chemistry, Delhi Technological University, Delhi 110042, India*

The emergence of textile-based wearable sensors as light-weight portable devices to monitor desired parameters, has recently gained much interest and has led to the development of flexible electronics on non-rigid substrates. The flexible biosensors may result in improved sports performance, to monitor the desired bodies for injuries, improved clinical diagnostics and monitor biological molecules and ions in biological fluids such as saliva, sweat. In addition, they could help users with different types of disorders such as blindness. In this context, new composite and nanomaterials have been found to be promising candidates to obtain improved performance of the textile based wearable devices and to optimize the structures for intimate contact with the skin for better functionality. This review aims to provide the most recent cutting-edge information on emergence, fabrication, materials, and applications of chemical and physical flexible and stretchable textile-based (bio)sensors. Besides this, we discusss the recent key innovations and applications of textile-based sensors in healthcare.
© 2020 The Author(s). Published on behalf of The Electrochemical Society by IOP Publishing Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 License (CC BY, http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse of the work in any medium, provided the original work is properly cited. [DOI: 10.1149/1945-7111/ab6827]

Manuscript submitted September 16, 2019; revised manuscript received December 24, 2019. Published January 21, 2020. *This paper is part of the JES Focus Issue on Sensor Reviews.*

The design and development of flexible and wearable sensor systems based on textile for measuring and quantifying physical and chemical signals generated by the human body provide an opportunity for disease diagnosis, therapy, security, robotics, and health monitoring.[1–6] Current advances in textile technologies, new materials, nanotechnology, and miniaturized electronics are making wearable systems more feasible.[7,8] Figure 1 shows different wearable sensors based on a textile that have attracted considerable commercial interest and vigorously expanded their scope for applications. For example, integrating sensing capabilities into textiles can provide significant benefits in diverse fields, such as healthcare, sports performance, and military issues.[3,9,10] These sensors hold considerable promise for maintaining and improving the quality of life while reducing medical costs. Also, these sensors can be used to remotely control personal health status as well as the physical state of persons (e.g an athlete, a soldier etc).

Recently, there has been an increased demand for wearable devices in the market, as demonstrated by the growth of the wearable fitness market to $34 billion in 2020 and this rate is expected to increase in future.[18,19] There are a variety of sensory systems that can be used in this area. It may be noted that most of these existing methods are complex and require sophisticated equipment that may not be readily available and applicable. Thus, the development of a new, small, safe, easy to use, sensitive and rapid method of detection still remains a challenge. Among all the sensors, some electrochemical devices have been developed that enable real-time monitoring both of chemical species in biological fluids (sweat, urine, tears), explosive residues and toxic gas,[20–23] and physical changes, vital human signs such as body temperature, for monitoring body position, temperature, stretching and movements.[3,10,24,25] Some techniques currently exist for the fabrication of electrochemical (bio) sensors and bio-devices based on textiles and several similar flexible platforms such as a shirt, bandages, glove and belt have been proposed (Fig. 1).

The advent of a new era of the flexible sensor will be expedited by innovations in flexible electronic and power sources.[26–29] Based on these findings, the new research direction of electronic flexible on textile has recently been launched. Also, more importantly, overcoming the design limitation, along with the fabrication of reliable electrochemical materials with optimized attributes and functionality, will facilitate the development of flexible and wearable devices. These smart devices can be miniaturized and equipped with remote and antenna systems for online monitoring and remote detection.[30,31]

Textiles represent an attractive class of substrates for fabricating wearable chemical smart sensors. The idea of wearable electronics or smart fabric will require the fabrication of devices directly into the fabric and these integrated electronic devices need to be flexible, light-weight and durable.[20,32] In addition, the challenge of the wearable sensor lies in linking the computer hardware to clothing or textile sensor. Steve Mann in the 1980 s, at Massachusetts Institute of Technology (MIT) implemented this technique by attaching various kinds of electronic devices to garments.[33] They used different methods to attach electrical components directly onto the fabric as shown in Fig. 2a user has to carry heavy rigid boxes and instruments and some wires in inflexible areas with visible wires through the cloths are impractical for daily use and cause users unsatisfied and uncomfortable. Currently, most of the research on wearable electronics is focused on the attachment or fabrication of electronic devices onto the textile for example with conductive yarns and using remote or Wi-Fi systems; microcomputers make them more exciting and convenient for end-users.[33,34] The intrinsic rough surface of textiles is problematic for electronic device fabrication. However, this property is ideal for many devices in which large surface roughness is actually preferred.

Conductive Materials and Fabrication Methods

One important consideration for wearable electronics is the choice of materials used in the fabrication and the possibility to provide good electrical performance with desirable mechanical properties, and stability into the flexible devices. The development of conductive and (semi)conductive materials, which are flexible, are necessary for wearable electronics. These smart materials are incorporated into the textile structure by different technologies

[z]E-mail: amirhatchem@yahoo.com; bansi.malhotra@gmail.com

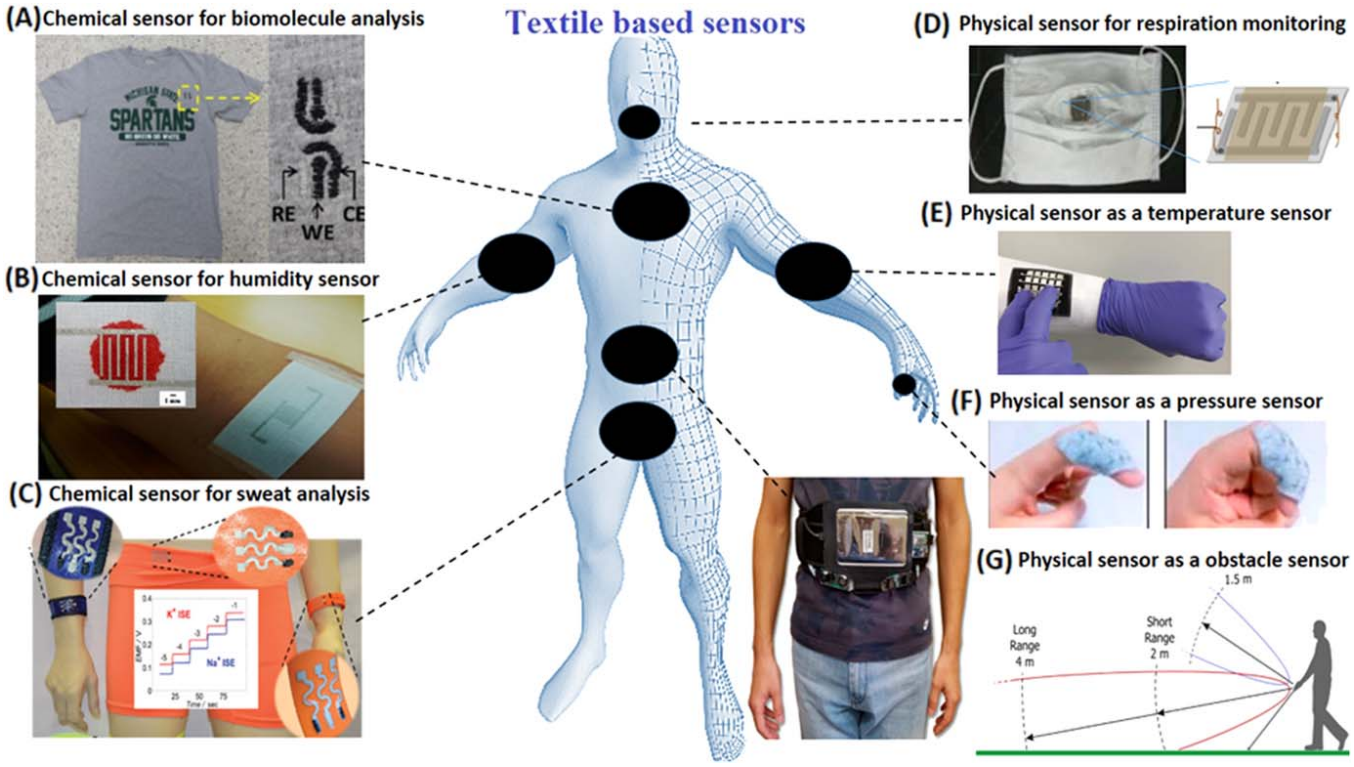


Figure 1. Two main applications of textile based sensors: Chemical sensors: (A. Biomolecule analysis,[11] B. Humidity sensor,[12] C. Sweat analysis,[13] and physical sensors: (D, Respiration rate sensor.[14] E, Temperature sensor,[15] F. Pressure sensor[16] G. Obstacle sensor.[17]

such as weaving, embroidering, knitting, printing, chemical treatments[32] and coating (coating the fibers with metals, galvanic substances or metallic salts by electroless plating and with a conductive polymers).[38,39] The different materials like conductive polymers (e.g polypyrrole, polyaniline),[38,40] metals and metal oxide nanoparticles/nanowires/nanorod (such as silver and zinc oxide),[39] carbon-based micro/nanomaterials (such as graphene, carbon particles, and carbon nanotubes),[40] have been used and investigated as textile modifier. Generally, in addition to metallic yarn and strain, smart materials can be grouped into three categories: conductive organic polymers, carbon-based micro/nanomaterials, and conductive metallic nanostructures. The following section discusses some of the materials and the fabrication technology frequently used to accomplish these goals.

Conducting polymers.—Conductive polymers are organic polymers exhibiting electrical properties similar to those of metals. Polypyrrole, polyaniline, and polythiophene fibers are good examples in practical flexible and stretchable electronic devices, and they can be produced with different sizes and even in nanometer dimensions, with high conductivity (120–130 S cm^{-1}) at room temperature.[10,41,42] To provide the necessary biocompatibility and increased resistance to mechanical stress these polymers can be modified and produced in different shapes such as fiber or nanofiber. Nanofiber is a class of nanomaterials having a diameter in the submicron range (50–500 nm), small pore size, sometimes with high porosity and high surface area with length is in the micron or millimeter range such as polyaniline nanofibers.[43] Due to their high surface area, they possess unique properties that make them potential candidates for a wide range of applications such as filtration, barrier fabrics, protective clothing, wipes and biomedical applications such as scaffolds for tissue engineering.[44] Different methods such spin coating, printing or chemical or electrochemical treatment can be used to synthesize polymer nanofibers.[45] Among these, the printing provides a convenient route to the deposition of conducting materials such as polymers with spatial resolution in the X,Y plane and makes a layer with thicknesses in nanometer dimensions. In addition, the screen-printing method provides an easily adopted fabrication route for the fabrication of wearable electronics, where all the layers with different functions can be printed on top of the fabric substrate with a layer-by-layer process. Another significant advantage of this method is that they facilitate the use of low-cost patterning techniques at room conditions.[46–48] Generally, the most promising materials are organic polymers because they are intrinsically flexible or can be easily used for making desired composite materials with special chemical, physical, and electronic properties.

Carbon-based micro/nano materials.—Zero, one and two dimensional carbon-based micro/nanomaterials such as graphene (nano nanosheets), carbon nanotube, carbon nanoparticles and its derivatives are promising materials for flexible and wearable electronics, due to their unique properties such as their high electron transfer rate and mechanical properties, large surface area, environmental stability and the potential for production at low cost.[49]

Among the various carbon materials, carbon nanotubes and graphene are two of the most intensively explored carbon allotropes in electrochemical sensors that have been evaluated as the electrode materials with great potential in wearable electronics.[50] A stretchable, porous, and conductive textile has recently been developed by a simple 'dipping and drying'' process using a carbon nanotubes ink.[51] Interestingly, some efforts have been made to assemble carbon nanotubes into macroscopic fibers in which the carbon nanotubes are highly aligned to maintain high mechanical strength and electrical conductivity of individual carbon nanotube. Particularly, carbon nanotube fibers are anticipated to have a broad range of potential applications, such as fiber-based microelectrodes and (bio) sensors.

Conductive metallic nanostructures.—Much attention has been paid towards the preparation and applications of nanostructured coatings onto textile substrate due to their promising applications.[52,53] Low dimensional nanostructures of metals, e.g., nanowires, nanorods (ZnO, Ag) or nanoparticles (such as Au, Ag, or Pt), are particularly attractive for fiber-based flexible and wearable electronics because some of these like gold and silver have very high

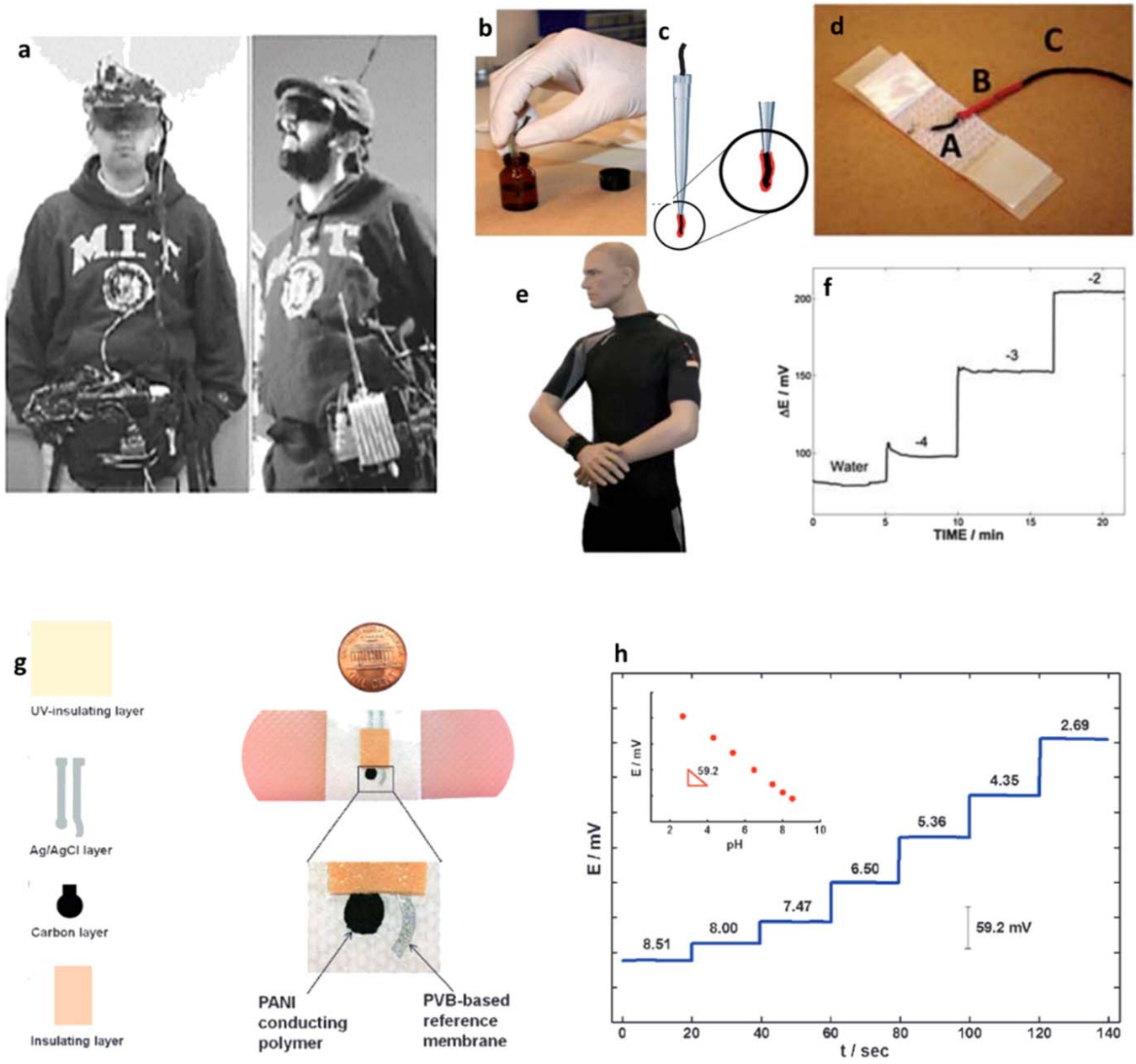


Figure 2. Recently developed wearable electrochemical sensors based on textile: (a) Wearable computing devices developed by Steve Mann (Reprinted with permission from Taylor & Francis, Copyrights 2016).[35] (b)–(f) Fabircaiton technique, schematic of prototype electrode and illustration of a K^+ sensing band-aid with the membrane-coated sensing section, respectively. The sensor shows (f) the response obtained when sensor is exposed to different concentration of K^+. (Reprinted with permission from the Royal Society of Chemistry, Copyright 2013)[36] (g) pH-sensitive bandage: Fabrication process and (h) potentiometric time-trace of a pH bandage sensor from pH 8.51 to 2.69 (Inset: EMF (mV) vs pH). (Reprinted with permission from WILEY-VCH, Copyright 2014).[37]

conductivities.[10,39,53–56] The nanostructured films deposited on textiles have exhibited excellent properties such as conductivity and antimicrobial activity. However, stability, is currently a big challenge and efforts are being made on improving the stability of metallic nanowires and nanoparticles on a textile substrate and utilizing them for inflexibly electronic applications.[52] Moreover, nanocomposite materials are good choice, as the polymers can be combined with nanostructures that may acquire sensing capability via conductive materials printed onto fibers or fabrics. The various approaches to make fiber-based flexible and wearable electronics by using conductive metal(s) can be grouped into two categories. In the first one, the electronic textile can be fabricated by using conducting fibers made from conductive metal such as steel. While the second category is thin-film-based electronic components where the different methods such as electro-spinning, wet spinning, spin coating, printing or chemical treatment (hydrothermal method) are being used for the preparation of materials. The printing provides a convenient route to the deposition of conducting nanocomposite materials with different thicknesses and the desired materials can be printed on top of the textile substrate through a layer-by-layer process.[10,46,53,57] In all cases, the optimal textile candidate for supporting printable electrochemical sensors have been predicted to have inert properties and yield stable operation for extended periods under normal and heavy wear.

Applications of Textile Sensors

Along with flexible electronic, textile electronics and its applications in flexible sensors are attractive for many applications. Textiles represent a striking class of substrates for fabrication of wearable

biosensors. Smart textiles describe the convergence of electronics and textiles into fabrics that can be used to sense, compute, communicate and actuate. Wearable biosensors ideally should have the ability to monitor the physiology of the wearer and the surrounding environment. The integration of chemical sensing into textiles thus permits the wearer to receive extremely valuable and timely information without compromising the functionality or comfort of the garment.[3,10,20,58–60]

The sensors can be divided into active and passive sensors. The passive sensors require an external power source, while active sensors can convert the input energy (elastic, thermal, etc) into a measurable difference of potential. The majority of textile sensors are known to be passive. Also, all of these systems require a power supply to function.[34,61,62] Because of the portable nature of these devices; batteries can be employed in almost most of them. At first batteries were unfixable. They could not be fully integrated with smart textiles. Developing flexible batteries is seen as an indispensable enabling technology for smart textiles and wearable sensors. The computers can be coupled to smartphone platforms directly and all wires can be eliminated.[63–65] We will discuss the recent advances and developments in the field of textile or wearable sensors with emphasis on two subclasses of these sensors i.e. chemical and physical wearable sensors.

Chemical wearable sensors based on textile.—The chemical wearable sensors include (bio) electrochemical and optical and gas sensors where a chemical reaction with the participation of the analyte (ions or (bio) molecules) gives rise to the analytical signal (Figs. 2 and 3). The gas sensors classified into chemical sensors are based on change in the conductometry. Wearable (bio) chemical systems can be used for health monitoring, military applications etc.[62] Among the different sensors, electrochemical (bio) sensors have a unique position for addressing the needs for field screening and identification of various chemical from ions to gas molecules. They have been found to be suitable for manufacture and miniaturization. Simplicity, low-voltage operation and low-power consumption of electrochemical sensors make them suitable for the requirements of on-body wearable systems into clothes. They are compatible with diverse materials and can operate with inexpensive battery-powered electronic instrumentation.[61,62] Electrochemical wearable sensors based on textile are being used for the detection of ions (Na^+, H^+, Ca^{2+}, NH_4^+, etc) (Figs. 2 and 3), biomolecules (glucose, lactate and uric acid) in different body fluids (sweat and saliva) and volatile compounds (such as: CH_4, NO_2, NH_3, CH_3OH, $(CH_3)_2CO$ and body odor) typically present in a body or in our environment (Table I).

In this kind of sensors, voltammetric, potentiometric, and amperometric techniques can be used. These sensors and their parts can be embedded into or on textiles such as in belt, arm hand, shirt, underwater garment, bandages and jacket and etc.[61,75,76] Numerous sensors based on textiles, cloths or yarns that show potential for integration with clothing can be used as (bio) sensor. Table I shows the characteristics and analytical performances of the textile electrochemical (bio) sensors. The researchers have tried to design the architecture of the removable sensor to eliminate problems commonly associated with the washing of integrated textile sensors. For fabricating clothes with electrochemical sensors, the direct screen-printing of electrodes on the textile substrate is convenient and straightforward technique.[20,46] However, it depends on differences in the hydrophobicity of various textile substrates such as cotton, polyester, and Gore-Tex (waterproof, breathable fabric).[20] Owing to the extreme hydrophobicity exhibited by the Gore-Tex fabric in comparison with other natural, synthetic fabrics, electrochemical sensing is facilitated; it has been shown to be a promising substrate for thick-film electrochemical sensors. Another attractive strategy was reported by J. Wang et al.[36] wherein the cotton yarns (Fig. 2b) were immobilized with carbon nanotube ink and it was used as a potentiometric sensor for detection of pH, K^+, and NH_4^+. Finally, they could fabricate a novel pH-sensitive bandage (Fig. 2g). The potentiometric time-trace of a pH bandage sensor from pH 8.51 to 2.69 along with EMF dependence vs pH for the PANI conducting polymer ISE is shown in Fig. 2h.

In sports (Figs. 3a–3f), bioelectrochemical[21,76,74] textile sensor plays a critical role. Biosensors are closely related to chemical sensors, distinguished by the molecular recognition element which is biological in nature (e.g. enzyme). The glucose[68] and lactate[74] sensors based on textiles have also been investigated. Figures 3a–3d are typical chronoamperograms of the tyrosinase-modified neoprene SPE exposed to catechols (such as 4-chlorophenol and phenol) and the stability of the signal over time with respect to the initial measurement for proposed sensors, respectively. As can be seen they developed flexible electrochemical sensors on underwater garments that are able to detect trace levels of marine pollutants and explosive threats in seawater. The electrode response stability was examined over 50 min, relative to the initial measurement for (i) 4-chlorophenol and (ii)phenol. These tests yielded a low RSD for each compound, ranging from 2.07% to 1.73%, respectively.

By using smart textile sensors, it is estimated that 70% of all illnesses are preventable, if suitable screening measurements could be introduced, this could produce dramatic reductions in costs for treatments and medication.[3,87] For example, by adopting suitable exercise and diets, many lifestyle-related illnesses can be prevented. Exercise generates sweat which contains rich information about the physiological condition of the subject as it contains a matrix of essential ions and molecules.[36,78] Real-time sweat analysis (Figs. 3f–3g) during exercise can give valuable information on dehydration and changes in the number of essential biomolecules and ions such as K^+ and Na^+ ions. Sweat, saliva, and tears contain multiple physiologically relevant chemical constituents and can be easily obtained continuously for non-invasive real-time monitoring of these analytes. For example, by monitoring level of electrolytes (pH, sodium, ammonium, calcium) and metabolites (glucose, lactate, urea) in sweat and the skin interstitial fluid with suitable sensors can provide valuable health information.[88,89]

Despite the extensive progress made in wearable electrochemical (bio)sensors for physiological monitoring, military application, and sports, little attention has been directed at the optical wearable (bio)sensors,[68,67] due to the difficulty for use in optical systems and related instruments into textiles and clothes. These sensors require the desired sample for analysis. Usually, a body fluid such as sweat is delivered to the sensor active surface, after which a reaction occurs, and a signal is generated in a few minutes in these sensors, textiles can act as a micro pump. For example, Schazmann et al.[88] reported a smart belt that electrode set on textile and it was located on the body to monitor Na^+ ions in sweat in real time. In this sensor, textile plays the role of sweat pump to the sensor area.

In the optical sensors, a colorimetric chip based on iono-gel (Fig. 4a) has been developed for monitoring pH of sweat.[68] There are many factors that can be correlated with sweat pH and health. Changes in the pH of the skin are reported to play a role in the pathogenesis of skin diseases like irritant contact dermatitis and acne, and also, showed that inducing metabolic alkalosis through the ingestion of sodium bicarbonate may lead to increased blood and sweat pH. In addition, it has been reported that sweat pH rises in response to an increased sweat rate. Schematic illustration of the micro-fluidic system's performance over and structure of the probe is shown in Figs. 4a–4c. Accordingly, The probe is applied to the monitoring of the pH of sweat during an exercise time. It was the first wearable sweat sensor with an ionogel being an important component in the fabrication of the sensing platform. The ionogel plymer provides an ideal substrate for the organic pH indicator dyes (bromophenol blue, bromocresol green, bromocresol purple and bromothymol blue). It could be attached onto the textile and could be used for monitoring the pH from 0 to 14 in intervals of one pH unit (Fig. 4c). This sensor showed no leaching of the indicators during experiments, the prolonged stability of the device for the monitoring of sweat pH measurements over time and the micro-fluidic structure (Fig. 4a) guarantees that freah sweat is continuously transferrd from the skin to the ionogels.

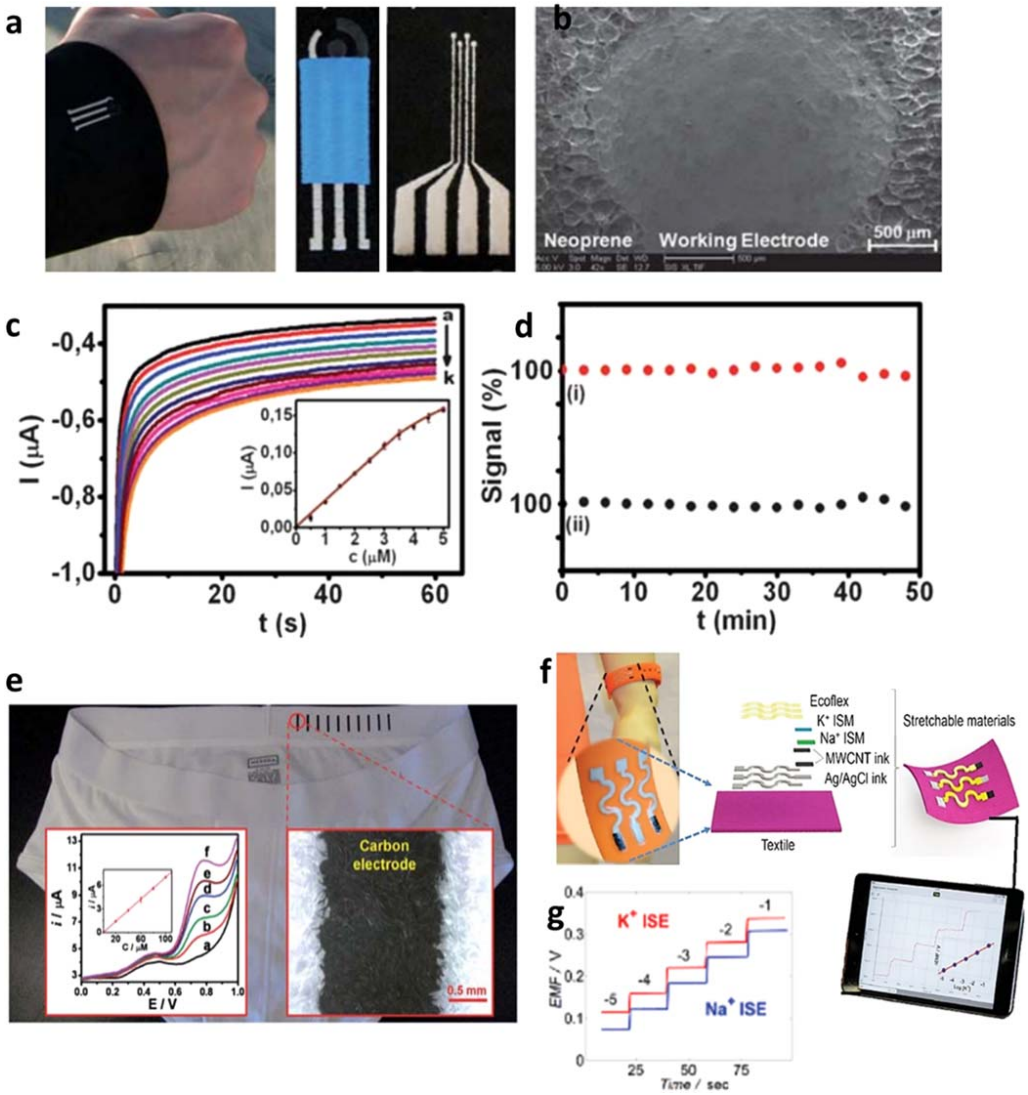


Figure 3. Recently developed wearable electrochemical sensors based on textile: (a) Wearable screen-printed electrochemical sensors on underwater garments and scanning electrode image (b) representing the working electrode area. (c) Chronoamperograms of the tyrosinase-modified neoprene SPE in the presence of different concentrations of catechol samples and corresponding calibration curves (inset). (d) Stability of the signal over time with respect to the initial measurement: (i) 10.0 mM 4-chlorophenol and (ii) 2.0 mM phenol (Reprinted with permission from Taylor & Francis, Copyrights 2016)[35] (e) Screen-printed carbon electrodes on the underwear. Single electrode morphology (right inset) linear-scan voltammetric response for increasing NADH concentrations over the 0.0–100.0 mM range (left inset) are also presented. (Reprinted with permission from the Royal Society of Chemistry, Copyright 2013).[36] (f) Screen-printed textile-based stretchable multi-ion (K^+ and Na^+) potentiometric sensor with its response(g) to different concetrations of K + (red) and Na + (blue) ions (EMF (mV)/time (sec). (Reprinted with permission from WILEY-VCH, Copyright 2014).[37]

Shirley Coyle et al.[66] reported a new optoelectronic probe on textile using a simple photometer (Fig. 4d) where a textile-based platform with fluid handling properties was used to collect and analyze the sweat samples. A fabricated channel is created by screen printing an hydrophobic acrylic paste on either side of the fabric. At one end of this channel is a window where sweat enters from the skin, based on capillary flow through the fabric. For sensing pH, a sensitive dye (bromocresol purple) is immobilized onto the channel (Fig. 4e). The dye for sweat analysis exhibits color changes from yellow to blue over this pH range A paired emitter-detector LED was used to quantify the color change of the dye. The fabricated optoelectronic textile based sensor was embedded into a waistband as a model and tested on the body during exercise succefully.

Recently, N. Promphet and et. al.[72] developed a textile-based colorimetric sensor for the detection of lactate and sweat pH and by depositing three different layers (chitosan, sodium carboxymethyl cellulose and indicator dye for pH sensing for lactate assay) onto a cotton fabric (Fig. 4g). The sensing probe for pH was composed of a mixture of two organic dyes (methyl orange and bromocresol green). As the pH of sweat from 1−14, the pH indicator shifted from red to blue, whereas the purple color intensity was enhanced by increasing lactate concentration (reange 0–25 mM) in the sweat (Figs. 4h–4i). The proposed sensor was tested on three human volunteers before and during jogging successfully.

Physical sensors based on textile.—The physical wearable sensors (temperature, pressure, obstacle and strain sensors) can be fabricated using textiles (Fig. 5). In general, a physical sensor, defined as a device that detects a change in a physical stimulus, turns that stimulus into a signal that can be measured or recorded.[34,78]

These sensors have drawn considerable attention, especially in medicine and sports. With the increase in the size of the elderly population, at the same time the emergence of chronic diseases because of the changes in lifestyle, there is a need to monitor the health status of individuals in their daily routine lives. The regular activities of the human body generate various physical stimuli, such

Table I. (Bio) electrochemical and optical and gas wearable sensors based on textile.

Sensor	Analyte	Clothing/the role of textile	Sample(s)	Sampling Method of sensor	Measurement technique	Sensing parameter	References
Optical	pH	clothing, head- or wristbands	Sweat	Textile channel	Colorimetric/LED/indicator	pH range: 2.0–8.0	66
Optical	H_2O_2, pH	—	—	Textile channel	Colorimetric	pH: 3.0–7.0 LDR: 0.1–0.6 μM	67
Optical	pH	waistband	Sweat	Textile channel	Colorimetric/indicator array	pH range: 4.5–8.0	68
Optical	pH	Silk fabric – Textile sensor	Polluted water	—	UV–vis reflectance absorption spectra	pH range: 2.0–10.0	69
Optical	Glucose/ Temperature	Cotton fabric- Shirt	Sweat	Textile channel	Colorimetric	LDR:50–600 μM LOD:13.45 μM	70
Optical	Levofloxacin	—	—	On Body	Surface-enhanced Raman spectroscopy	N.R.	71
Optical	pH and lactate	Cotton fabric	Sweat	On Body	Colorimetric	pH: 1.0–14.0 LDR:0.0–25.0 mM	72
Optical	Ammonia and Hydrogen chloride	Cotton Fabric treated with chemical indicator	Air	—	Image analysis	LDR: 50–1000 ppm	73
Electrochemical	Lactate	woven electrode into textile	Sweat	On Body	Amperometric	LDR: 4.0–-20.0 μM	74
Electrochemical	Blood leakage sensor	Cotton fabric.	Blood	On Body	Impedance	LVOD: 15 μl	12
Electrochemical	Nitro-aromatic Explosives	screen-printed electrodes on textile	Air	—	Voltammetry	LDR: 0.0–50 ppm	20
Electrochemical	lactate	Cotton fabric.	Salvia	—	Amperometric	LOD: 1.0 ppm LDR: 0.1–5.0 mM LOD: 0.3 mM	21
Electrochemical	Phenolic contaminants, TNT, Cu^{2+}	Nitrile glove underwater garments	Marine environments	On Body	Amperometric, voltammetry	[TNT]: LDR:100 – 900 ppb LOD: 42 ppb [Cu^{2+}]: LDR:10–90 ppb LOD: 13 ppb [4-chlorophenol]: 0.0–25.0 μM LOD: 0.43 mM [Phenol]: LDR: 1.0–5.5 μM LOD: 0.25 mM	75

Table I. (Continued).

Sensor	Analyte	Clothing/the role of textile	Sample(s)	Sampling Method of sensor	Measurement technique	Sensing parameter	References
Electrochemical	NADH and H_2O_2	screen-printed electrode on the underwear	Sweat, and Salvia	On Body	Amperometric	NADH: LDR: 0–100 mM LOD: N.R. H_2O_2 LDR: 0–25 mM LOD: N.R.	76
Electrochemical	Na^+, K^+, Cl^-,Cd^{2+}, Pb^{2+}, pH	textile-based sampling	—	Textile channel	Potentiometric	[Ions]: LDR: 10^{-7}–10^{-1} M	77
Electrochemical	pH, K^+, NH_4^+	band-aid/conductive yarns coated with polymeric membrane	Sweat	On Body	Potentiometric	pH: 3.0–11.0 K^+ LDR: 0.01–100 mM LOD: 10.0 μM NH_4^+ LDR: 0.001–10 mM LOD: 1.0 μM	36
Electrochemical	pH	Bandages/screen-printed pH potentiometric sensor	Sweat	On Body	Potentiometric	pH: 3.0–10.0	78
Electrochemical	Na^+,K^+	screen-printed potentiometric sensor	Sweat	On Body	Potentiometric	[Na^+] and [K^+]: LDR: 0.1–100 mM LOD: 10–4.9	13
Electrochemical	Glucose or lactate $^-$	Polyester(T- Shirt)	Sweat	On Body	Amperometric	Glucose LDR: 0.0–40 mM LOD: N.R. Lactate LDR: 0.0–50 mM LOD: N.R.	11
Electrochemical	Uric acid	Bandage/ screen-printed potentiometric sensor/ wireless	Sweat	On Body	Amperometric	LDR: 100.0–800 μM LOD: N.R	22
Electrochemical	Adrenaline	Cotton yarn functionalised with Organic electrochemical transistors	Sweat	On Body	Oxidation of adrenaline at the Pt-gate electrode	LDR: 0.001–1.0 mM LOD: N.R	79
Electrochemical	Cl^-	polyester	Sweat	On Body	Potentiometric	LDR: 10.0–100 mM LOD: N.R	80

Table I. (Continued).

Sensor	Analyte	Clothing/the role of textile	Sample(s)	Sampling Method of sensor	Measurement technique	Sensing parameter	References
Electrochemical	Na+	carbon fibres integrated into wearable patch, garments	Sweat	On Body	Potentiometric	LDR: 1.0–100.0 mM LOD: N.R	81
Electrochemical	Interleukin-6	Textile coated with layers of silver and carbon	blood	—	Stochastic microsensors	LOD:1.0 fg. mL^{-1}	82
Electrochemical	DMP, NH_3, NO_2	polymers coated Nylon	Air	—	Conductometry	N.R.	83
Electrochemical	NH_3,HCI	PPy-coated textile fabrics	Air	—	Conductometry	N.R.	84
Electrochemical	Solvent vapors	Carbon black/polymr gas sensors integrated into a textile	Air	—	Conductometry	Acetone LDR: 50.0–500 ppm LOD: N.R	85
Electrochemical	NH_3	Modified commercial yarns with polyaniline	Air	—	Conductometry	—	11
Electrochemical	influenza A virus	A textile screen-printed biosensor on polyamide substrates	Biofluid	On body	Potentiometric	LDR: 10 ng ml^{-1} to 10 μg/ml LOD: 10 ng ml^{-1}	86
Electrochemical	Volatile amine	Nano-polymeric composite	Body odor	On body	Conductometry	LDR:50–1000 ppm	23

Abbreviations: [TNT]: 2,4,6-trinitrotoluene,[DMP]: Dimethyl methylphosphonate, [PPy] Polypyrrole, [LDR]: Linear dynamic range, [LOD]: limit of detection, [NADH]: Nicotinamide adenine dinucleotide, [N.R].: Not reported.

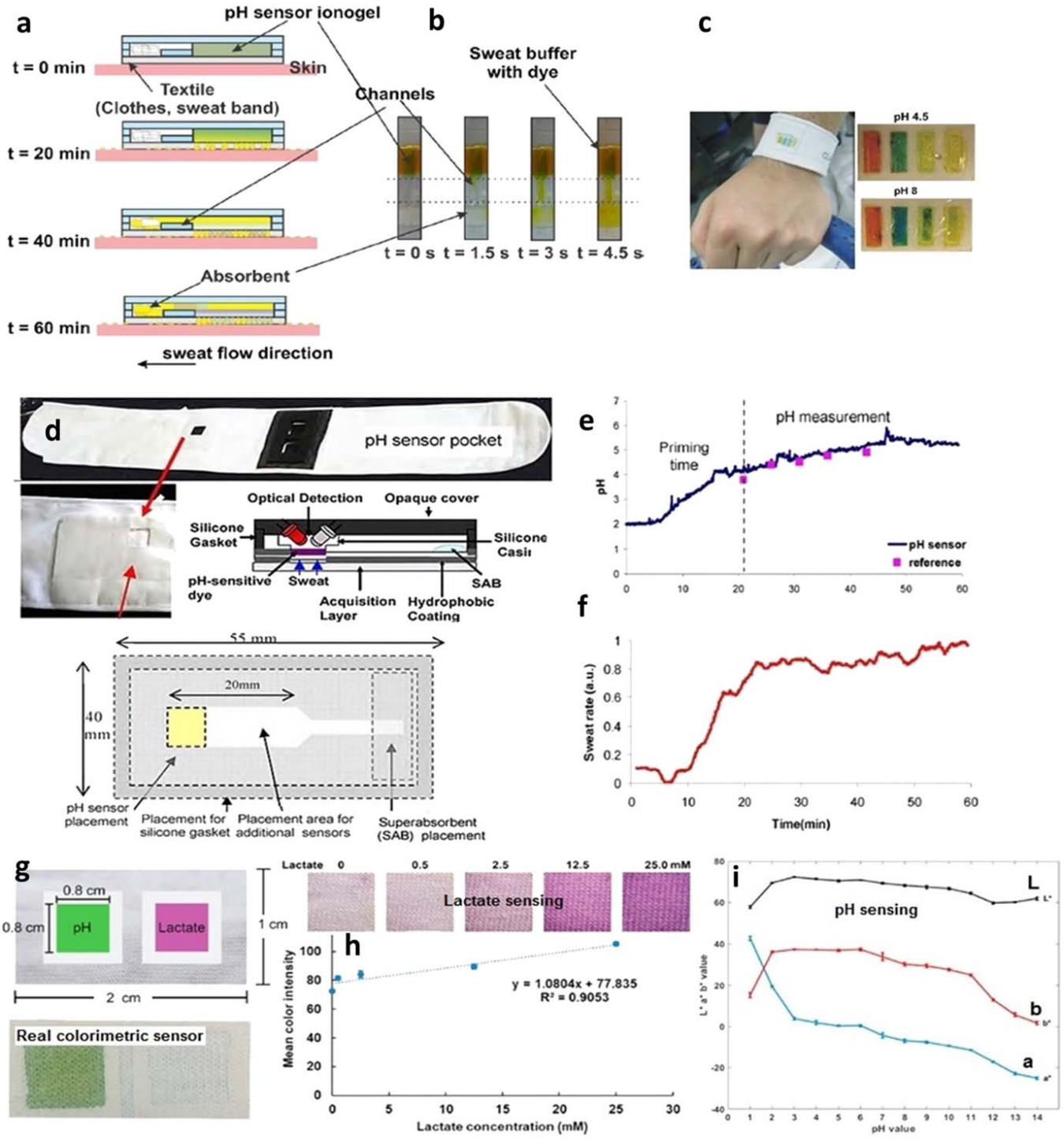


Figure 4. Recently developed wearable optical and optoelectronic sensors based on textile: (a), (b) Schematic illustration of the micro-fluidic system's performance of wearable sweat sensor over time c) Microfluidic system integrated into a wrist-band (Reprinted with permission from Elsevier, Copyright 2012).[68] (d) pH sensing patch (top) and fabric passive pump design (down) (e) pH and (f) sweat rates measured during a 60 min effort of two volunteer (Reprinted with permission from IEEE, Copyright 2009).[66] (g) shematic representation of a colorimetric sensor with (h) a standard color charts and real-time sweat pH and Lactate monitoring based on a textile platform. (i). Standard color charts of the textile based pH sensor (a) and L* a* b* color model of pH in 0.1 M PBS solution (b). (Reprinted with permission from Elsevier, Copyright 2019).[72]

as body temperature, blood pressure, the pulse skin and strain. On the other hand, physical sensors are capable of measuring these types of physical parameters or health indicators.[34] Generally, the four main vital signs routinely monitored by medical professionals are:

i. Body temperature (exploration of infection and fever).
ii. Heart rate (exploration of cardiovascular diseases).
iii. The rate of breathing or respiration rate (it may increase with fever, illness, and with other medical conditions, the typical respiratory rate for a healthy adult at rest is 12–20 breaths per minute).
iv. Blood pressure (in the exploration of hypertension, note: blood pressure is not considered a vital sign, but is often measured along with the vital signs).[10,34,78]

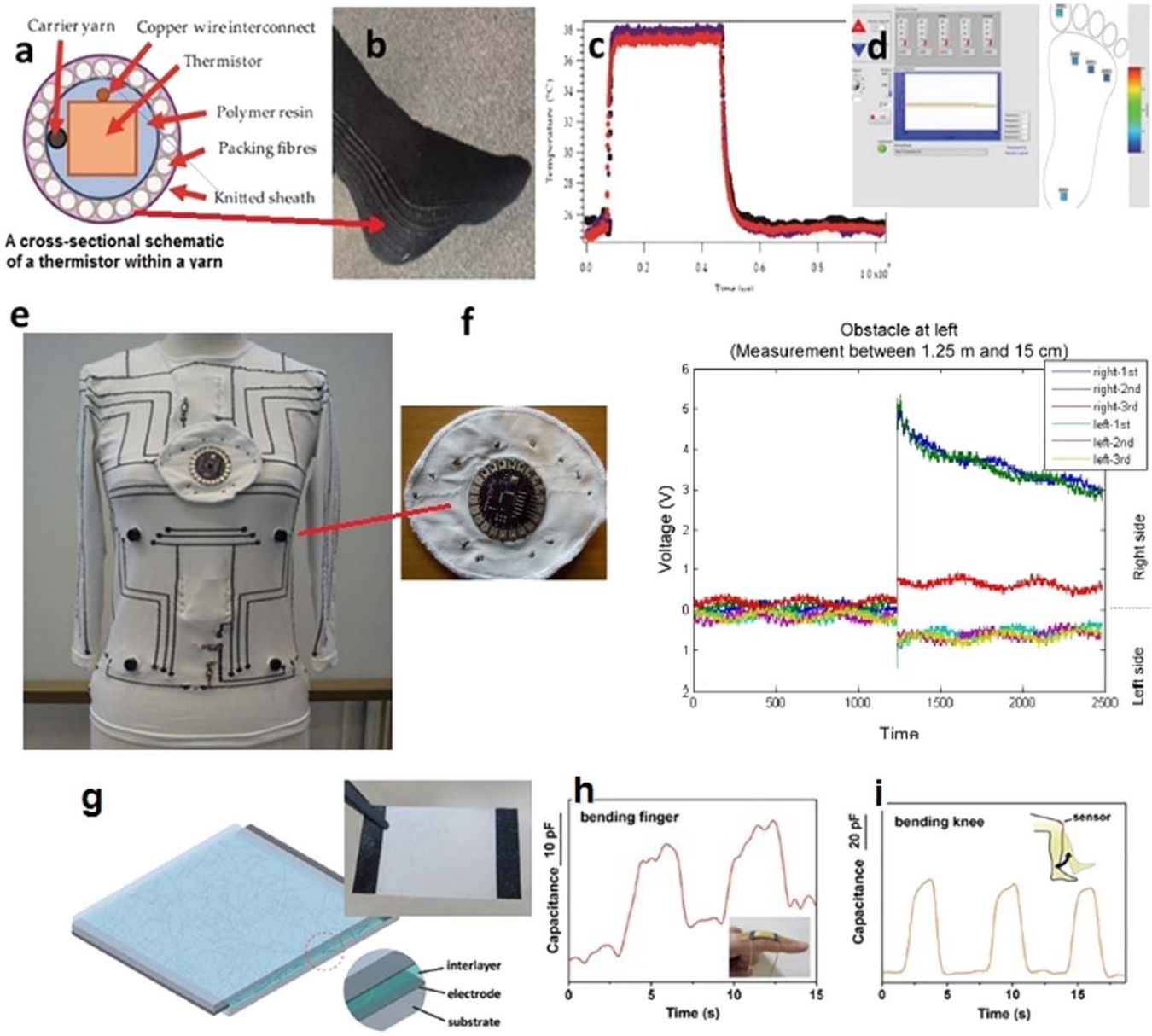


Figure 5. Recently developed wearable physical sensors based on textile: (a) An schematic of the encapsulation for a thermistor within a yarn and photograph of prototype sock (b) fabricated with electronic temperature-sensing yarns. (c) temperature response of three different sensors and Computer interface (d) using LabVIEW (Reprinted with permission from MDPI, Copyright 2017).[90] (e) Wearable obstacle detection system fully integrated into textile structures worn by the mannequin moving towards an object (f) recorded signals over the 1st, 2nd and 3rd vibration motors on the left and right arms during the movement towards front obstacle (Reprinted with permission from Elsevier, Copyright 2012).[91] (g) The schematic of the pressure sensor integrated into the bandage as a substrate for detection of articular and muscular movement: finger (h) and knee(i) bending.[24]

Researchers have fabricated a smart T-shirt that functions to monitor the wearer's heart rate, respiration, temperature, and other vital signs.[9] The Wearable Health Care System project has developed a wearable garment that can be worn under regular clothing and is capable of recording biomechanical variables and physiological signals.[33,92,93] In addition, motion signals were monitored because they can be used to extract important physiological information such as the stress level or movement of a person (in some research such as research on Parkinson's disease is very useful). In all of them textile and clothes can play critical roles, and body parameters can be detected with physical wearable sensors based on textiles.

Temperature sensor.—The body temperature is regarded as the first vital sign. An increase and decrease of body temperature can be due to an infection or fever and low blood flow rate (due to circulatory shock), respectively. Generally, body temperature is about 32 °C, and for measuring the temperature, the suitable places are the arm or the chest.[3,7] The textile-based sensor can be used as a temperature probe based on thermistors (in thermistor electrical resistivity changes with variation of temperature). Generally, in these sensors the following formula has been used.[94,95]

$$\ln R_t = \ln R_0 + \beta\left(\frac{1}{T} - \frac{1}{T_0}\right) \quad [1]$$

In this equation, a linear relationship between ln (R_t) and 1/T is established. Here, β (in Kelvin) is the material constant for the thermistor (it can be a metal oxide, carbon nanotubes, graphene, nickel oxide).[95–98] R_t is the resistance at temperature T, *and* R_0 is the resistance at T_0. β with a temperature coefficient of the thermistor have main roles in determining the sensitivity of the thermistor.[95]

The thermistors should have high static and dynamic accuracy in the human skin temperature range (30 °C–45 °C) and can be embedded into or attached on textiles. Detection should be linear with possibly high-temperature efficiency and resistant to humidity is expected as an environmental challenge.[99,100] Jung et al. recently found that combination of poly (3,4-ethylene dioxythiophene) poly (styrene sulfonate) (PEDOT:PSS), silver nanoparticles (AgNPs), and graphene inks can be used to fabricate a highly stretchable self-powered temperature sensor. Also, the printing technique with commercialized thermoelectric materials was utilized on stretchable knitted fabric with excellent elasticity as an affordable and scalable process. In fact, temperature gradients over the printable maters can be fabricated to control the current via the 'Seebeck effect''.[101] Hence, temperature difference across the device is detected automatically and without need to any power supplier.[15] In another approach, T. Hughes-Riley et al.[90] reported a new electronic temperature-sensing yarn that was well-suited for making long-term monitoring of skin temperature when incorporated into a textile with a remote system (Fig. 5a). The miniatured sensor were suitable for wearable applications. They implanted a commercial thermistor chip into the fibers of a yarn, which can be used to produce textile. To protect the thermistor, it was encapsulated in a resin. The fabricated textile showed a short response time over the temperature range of interest (between 0.01–0.35 s). Finally, they could make socks (Figs. 5a–5d) as a prototype temperature-sensor and it was suitable for diabetic patients.

Obstacle detection system.—The detection of obstacle in an environment of person is highly desired, in order to detect obstacles, different systems such as Radio Frequency Identification (RFI), Navigation Assistant for Visually Impaired (which was designed to convert images captured by a camera into verbal messages), GPS (Global Positioning Systems), Robotic, etc, can be used. These sensors can guide the users to detect obstacles such as a table, stairs, and even a hole.[91,102,103] Bahadir et al.[91] developed an innovative wearable obstacle detection system fully integrated to textile structures (Fig. 5e–5f), that enabled detection of obstacles for visually impaired people. This smart clothing included ultrasonic sensors, vibration motors (Fig. 4e), power supplies and a micro-controller. The prototype (Fig. 4e) has been tested for detection capability, power consumption, heating behavior, and wash-ability. Results have shown that developed system is able to identify the obstacle's position without any failure within its detection range around its users. Interstingly, the developed system is capable of detecting right, left and front obstacles position accurately and giving fast signal. While detecting obstacles such that when obstacles are at the left, system gives an output turn right and reverse and when obstacle is at the front, the smart cloth gives an output like turn left and right at the same time, so user can recognize there is an obstacle in front of the user. For example, Fig. 4f shows signals of vibration motors (1st, 2nd and 3rd) on the left and right arms during the movement towards an obstacle at distance about 15 cm to 1.25 m located at the left side.

Mocanu et al.[17] designed and fabricated a flexible textile-based obstacle detector using paper to attain a promising wearable device for both indoor and outdoor activities. In fabricated device, two independent sections of ultrasonic sensor and internal video camera were utilized to identify both static and mobile substances, irrespective of their dimension, shape or location and interpret and notify users about coming obstacles.[17]

Pressure sensor.—A textile-based pressure sensor has been explored for a variety of applications to include sports, caring for the elderly, diagnostics and human motion detection due to their integrability into clothes, socks, shoes, pillow, belt, headband, and even beds[102,104] (Fig. 4c). In all these sensors, the determination of pressure regions or sensor positions for example in health or sports cases are very important. As we know under the human feet or shoes, different pressure regions are available.

For measuring pressure, many types of sensors such as capacitive,[105] piezoresistive,[106] piezoelectric,[107] and optical[108] have been developed. Among these, a capacitive pressure sensor has advantages in terms of its simple design, low power consumption, acceptable sensitivity, and high stability. A capacitive pressure sensor makes use of the principle of a parallel plate capacitor, which consists of two compliant conductive layers (such as conductive fabrics) and flexible spacers (flexible dielectric such as foams, fabric spacers and soft polymers). The amount of electrical charge (Q) that can be stored between the two conductive plates of such a capacitor at a given voltage V is defined by the capacitance C of the capacitor. C can be expressed as a function of the distance d, the given area A and the material between the two plates, therefore, C can be defined as:

$$C = \frac{\varepsilon_0 \, \varepsilon_r \, A}{d} \qquad [2]$$

where, ε_0 ($\varepsilon_0 = 8.854 \times 10^{-12}\ \mathrm{F \cdot m^{-1}}$) is the vacuum permittivity and ε_r is the relative dielectric constant of the dielectric material. In order to improve the performance and efficacy of the capacitive sensor Many efforts have been made for the selection of a suitable interlayer material or modification with nano-conductive materials (metallic nanoparticles, such as silver nanoparticles as filler or doping agent), and enlarge the deformation space and extend the surface area of the electrodes.[109] A textile capacitor has been fabricated using two electrodes of conductive woven textile separated by compressible spacer material. Meyer,[110] developed a pressure sensing pillow with several integrated textile capacitive pressure sensors made from a 3D knit fabric spacer material and conductive textile filaments coated with silver. The developed sensor could be used to detect pressure with pressures up to 10 $\mathrm{N\,cm^{-2}}$. An ideal pressure sensor must ensure repeatable measures over time with a low hysteresis value. Therefore, the textile pressure sensors are applicable in areas where high accuracy is not required.[110]

In a similar work, a smart adhesive bandage[24] for monitoring human movement was developed (Figs. 4g–4i). For the fabrication of electrode of the pressure sensor, silver nanowire mixure dispersed in ethanol was drop-coated on to a polyethylene terephthalate film substrate to produce conductive networks. To monitor human movement, the PET film substrate was replaced by an adhesive bandage. Figure 4g shows the schematic of pressure sensor with low weight and high flexibility. The final sensor as a dielectric interlayer demonstrates high sensitivity, as 5.54 kPa−1 ($p < 30$ Pa) and 0.88 kPa−1 ($p > 30$ Pa), The adhesive bandage sensor was used as a substrate for detection of articular and muscular movement such as finger and knee bending (Figs. 4h–4i).

Heart rate (pulse) sensor.—Traditionally, heart rate can be measured from the radial artery at the wrist and monitoring of blood pressure is performed in a clinic with trained personnel by mounting inflatable pressure cuffs with stethoscopes to the patient's arm, the so-called auscultatory method.[7,10] Heart rate depends on cardithe ac cycle (the cycle is the sequence of events that occurs when the heart beats that deoxygenated blood through the lungs and pumping newly oxygenated blood). The heart rate is the frequency of cardiac cycles, expressed as beats per minute (b.p.m.). To measure the heart rate more accurately, electrical sensors can be used. Generally, skin electrodes are used to pick up the depolarization signal from the heart muscles. This technique is called as electrocardiography.[111,112] A few sensors based on textile have been reported although a lot of research-based on wearaa ble sensor was reported,[99,100,113]

Yoo and coworkers(100) designed a wearable electrocardiogram acquisition system implemented with planar-fashionable circuit board on a shirt. The proposed wearable continuous electrocardiography monitoring system is composed of two parts: a group of electrodes and a monitoring chip. The electrodes werescreen-printed directly on fabric. Unlike conventional wet electrode, to overcome these drawbacks of conventional wet electrodes (dry out of

electrolyte gel, unpleasant using of gel for a long time and also reduce of toxicity of gel) for long-term monitoring, dry electrodes can be considered as a good alternativece to the wet electrode.

Lee et al.[113] developed a non-contact electrocardiogram measurement system on a bed using conductive textiles for long-term monitoring of heart activities. At first, their goal was to design and develop more user-friendly system by configuring the array and size of electrodes with no constraints of posture and position in a bed, for this purpose they used a foam layer to provide comfort and follow the curvature of the body. Their results suggested that the capacitive coupled textile system was suitable for daily healthcare monitoring and abnormal state of sleeping. Finally, they tested the system to find the potential for monitoring the sleep stages, which is essential for observing the physical and psychological states.

To obtain a continues monitoring of electrocardiogram (ECG) for possible risk of cardiac arrest in babies, Coosemans et al. (114) developed an alternative to conventional ECG electrodes to be used in the long term with no skin irritation and allergic contact reactions. Figures 6a–6c shows the integrated circuitry and secondary coil and a sample set-up with the belt prototype on a three months old baby. General system overview is depicted in Fig. 6a. The measured ECG results of new electrode called Textrode demonstrated the lower frequent drift according to baseline along with a lower signal-to-noise ratio in comparison to conventional gel electrodes. Also, the reported heart rate accuracy is a beat-to-beat detection which suffices the application.[114]

Respiration rate sensor.—The respiratory rate is a vital sign with an under appreciated significance that can, in acute situations, prognosticate patients' mortality rate and need for invasive ventilation. There are many approaches for respiration monitoring, common sensors can respond to the flow of breath, but in physical sensors for respiratory rate based on textile, to the expansion and contraction of the chest during breathing can be monitored. The strain or pressure sensing can be accomplished by monitoring the volume changes of the chest.[3,30,34,116]

The respiration rate sensors take advantage of conductive textile technology to build the sensing elements directly into the textile.[116,117] Huang et al.[117] knitted nonconductive fibers with conductive silver-coated to form resistive strain sensing textiles and used it as reparation rate sensor. As the textile is stretched, the contact area between the conductive fibers is reduced, and the resistance increases.

Multiple textile-based pressure sensors (capacitive sensor) have also been reported. Hoffmann et al.[118] and Min et al.[119] fabricated a compressible textile wherein the foam material was sandwiched between two conductive textile electrodes. In another work, they simplified structural textile capacitive respiration sensor for respiration monitoring system. In this sensor whenever the chest expands, the pressure applied on the compressible textile reduces its thickness, thus increasing the capacitance between the electrodes.

Importance of sensor comfortability for user was also emphasized by Guay et al.[115] by proposing a prototype wearable designed

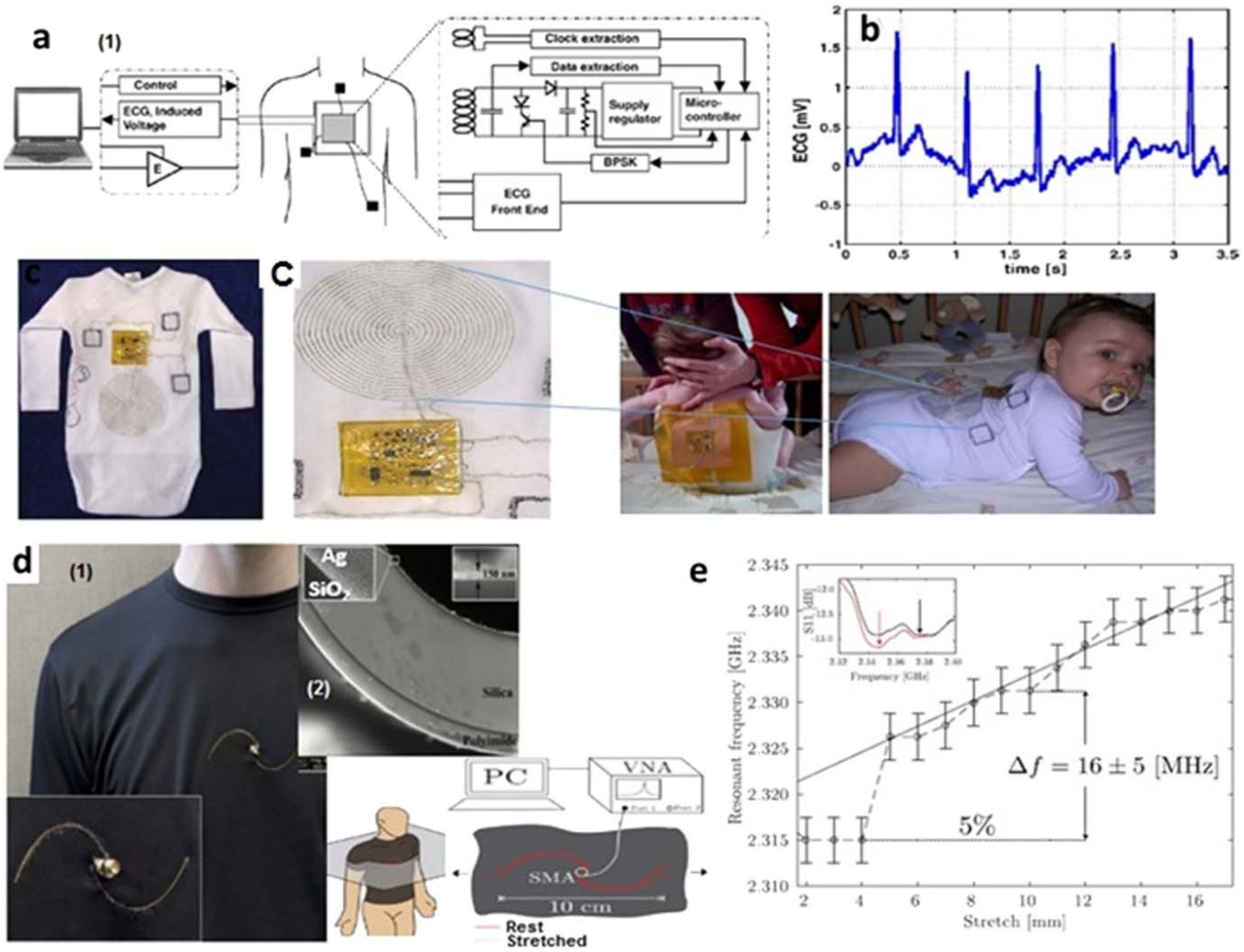


Figure 6. Textile based sensors: heart rate sensor (a) general sensor performance with correspond (b) ECG measurment (c) Detail of the backside of the baby suit, showing the embroidered coil and the flexprint (114). T-shirt with respiration rate sensor: (d) Prototype of the spiral antenna integrated into a cotton shirt supported by SEM image (top right) (e) Resonant frequency shift of the as a function of the induced stretch in off-body scenario.[115]

sensor to monitor the breathing rate without any direct skin interaction or add-on electrodes (Figs. 6d–6e). Their design was multi material fibers with form of spiral antenna detecting shift of central frequency coming from the lung volume variation and strechability of chest textile, which remotely records the respiratory information in real-time.[115] They reported that bearthing detection can be either due to changing the antenna configuration as result of mechanical force or changing the dielectric of human torso during respiration. The measured frequency shift for deep and shallow breathing is located in the ranges of 120–200 MHz and 10–15 MHz, respectively. As can be seen in Fig. 6d, linear line of 5–16 mm stretch range provides the respiration sensor sensitivity of 1.4 MHz mm^{-1}, and 5% chest expansion causes 16.2 resonant frequency shift.

Conclusion

The emergence of wearable textile-based sensors has opened a brighter future to a wide range of healthcare applications. In the last decades, the considerable progress in textile materials and fabrication techniques of sensors and flexible devices have revealed the importance of these sensing platforms. The wearable textiles-based sensors can be integrated with daily clothing However, there are still exist many challenges to be addressed in terms of fabrication, design, device operation and implementation. Despite the huge progress in 'engineering'' part of wearable snesors, many key challenges such as quantification analysis and stability need to be optimized. For example, the electrochemical methods need to be improved since the obtained signal is required to be compared with either a calibration curve or a standard method. Another serious issue is the stability of the wearable sensors particularly electrochemical sensors due to passivation and biofouling of the electrode surface, which possibly increases the risk of wrong response during operation. This problem becomes more important when the enzymes or biological recognition agents are used, especially during on-body measurements. Additionally, challenges relared to power sources still remain for these types of sensors. In fact, an ideal power source should have a small size and produce stable and enough energy, flexible to increase comfort and should be close to the sensor. So more robust and stable sensing mechanisms and components must be designed and should be used to enable accurate measurements for increased chemical and mechanical stability and long-term use. In addition, although nanomaterials have been found to play an important function in flexible sensors for enhanced sensing performance, their potential towards biocompatibility and toxicity are needed to be considered and investigated.

Addressing these problems and challenges in these areas are critical for the future success of the wearable sensors. It should be noted that most sensors are used for continuous monitoring of parameters related to health and user safety.Efforts should be made to explore the applications of these sensing devices in forensic areas and drug toxicology that are still underexplored. Indeed, we predict exciting development towards the understanding of operational flexible textile-based sensors platforms for new applications. Finally, we hope that the revolution in textile sensors will lead to produce new products that can be released in the markets for our improved daily lives.

Acknowledgments

Amir Hatamie and Abdolreza Simchi are thankful for the financial support from Sharif Universityof Technology (Grant QA970816) and Iran National Science Foundation (INSF Grant 95-S-48740). Also, the authors thank Prof. Yogesh Singh, Vice Chancellor, Delhi Technological University (DTU), Delhi, India for his interest in this work. SK (DST/INSPIRE/04/2017/002750) and CMP (DST/INSPIRE/04/2015/000932) acknowledges Department of Science and Technology, New Delhi, India for DST-INSPIRE Faculty Award. B.D.M. thanks Science & Engineering Research Board (Govt. of India) for the award of a Distinguished Fellowship (SB/DF/011/2019).

References

1. Y. Yu, H. Y. Y. Nyein, W. Gao, and A. Javey, *Adv. Mater.*, 1902083 (2019).
2. D. R. Agrawal, Y. Tanabe, D. Weng, A. Ma, S. Hsu, S.-Y. Liao, Z. Zhen, Z.-Y. Zhu, C. Sun, and Z. Dong, *Nature Biomedical Engineering*, **1**, 0043 (2017).
3. Y. Khan, A. E. Ostfeld, C. M. Lochner, A. Pierre, and A. C. Arias, *Adv. Mater.*, **28** (22), 4373-4395 (2016).
4. A. J. Golparvar and M. K. Yapici, *J. Electrochem. Soc.*, **166**, B3184 (2019).
5. D. Chung and B. Gray, *J. Electrochem. Soc.*, **166**, B3116 (2019).
6. P. Bhushan, Y. Umasankar, S. RoyChoudhury, P. A. Hirt, F. E. MacQuhaec, L. J. Borda, H. A. Lev-Tov, R. S. Kirsner, and S. Bhansali, *J. Electrochem. Soc.*, **166**, B830 (2019).
7. P. Rai, S. Oh, P. Shyamkumar, M. Ramasamy, R. E. Harbaugh, and V. K. Varadan, *J. Electrochem. Soc.*, **161**, B3116 (2014).
8. J. Jang and J. I. Han, *J. Electrochem. Soc.*, **164**, B136 (2017).
9. H. Lee, T. K. Choi, Y. B. Lee, H. R. Cho, R. Ghaffari, L. Wang, H. J. Choi, T. D. Chung, N. Lu, and T. Hyeon, *Nat. Nanotechnol.*, **11**, 566 (2016).
10. W. Zeng, L. Shu, Q. Li, S. Chen, F. Wang, and X. M. Tao, *Adv. Mater.*, **26**, 5310 (2014).
11. X. Liu and P. B. Lillehoj, *Lab Chip*, **16**, 2093 (2016).
12. K.-i. Nomura, Y. Horii, S. Kanazawa, Y. Kusaka, and H. Ushijima, *Sensors*, **18**, 240 (2018).
13. M. Parrilla, R. Cánovas, I. Jeerapan, F. J. Andrade, and J. Wang, *Adv. Healthcare Mater.*, **5**, 996 (2016).
14. B. Li, G. Xiao, F. Liu, Y. Qiao, C. M. Li, and Z. Lu, *J. Mater. Chem. C*, **6**, 4549 (2018).
15. M. Jung, S. Jeon, and J. Bae, *RSC Adv.*, **8**, 39992 (2018).
16. S. Y. Kim, E. Jee, J. S. Kim, and D. H. Kim, *RSC Adv.*, **7**, 23820 (2017).
17. B. Mocanu, R. Tapu, and T. Zaharia, *Sensors*, **16**, 1807 (2016).
18. H. Karvonen, C. Pomalaza-Ráez, K. Mikhaylov, M. Hämäläinen, and J. Iinatti, in *Advances in Body Area Networks I* (Springer, Berlin) p. 235 (2019).
19. R. Vooris, M. Blaszka, and S. Purrington, *Int. J. Sociol. Leisure*, **2**(4) 421 (2019).
20. M. C. Chuang, J. R. Windmiller, P. Santhosh, G. V. Ramírez, M. Galik, T. Y. Chou, and J. Wang, *Electroanalysis*, **22**, 2511 (2010).
21. R. S. Malon, K. Chua, D. H. Wicaksono, and E. P. Córcoles, *Analyst*, **139**, 3009 (2014).
22. P. Kassal, J. Kim, R. Kumar, W. R. de Araujo, I. M. Steinberg, M. D. Steinberg, and J. Wang, *Electrochem. Commun.*, **56**, 6 (2015).
23. T. Seesaard, P. Lorwongtragool, and T. Kerdcharoen, *Sensors*, **15**, 1885 (2015).
24. J. Wang, J. Jiu, M. Nogi, T. Sugahara, S. Nagao, H. Koga, P. He, and K. Suganuma, *Nanoscale*, **7**, 2926 (2015).
25. D. Vilela, A. Romeo, and S. Sánchez, *Lab Chip*, **16**, 402 (2016).
26. S. Huang, Y. Liu, Y. Zhao, Z. Ren, and C. F. Guo, *Adv. Funct. Mater.*, **29**, 1805924 (2019).
27. B. Sun, Y.-Z. Long, Z.-J. Chen, S.-L. Liu, H.-D. Zhang, J.-C. Zhang, and W.-P. Han, *J. Mater. Chem. C*, **2**, 1209 (2014).
28. X. Wang and G. Shi, *Energy & Environmental Science*, **8**, 790 (2015).
29. A. J. Bandodkar, W. Jia, and J. Wang, *Electroanalysis*, **27**, 562 (2015).
30. M. Catrysse, R. Puers, C. Hertleer, L. Van Langenhove, H. Van Egmond, and D. Matthys, *Sensors Actuators A*, **114**, 302 (2004).
31. G. Liu, C. Ho, N. Slappey, Z. Zhou, S. Snelgrove, M. Brown, A. Grabinski, X. Guo, Y. Chen, and K. Miller, *Sensors Actuators B*, **227**, 35 (2016).
32. M. Stoppa and A. Chiolerio, *Sensors*, **14**, 11957 (2014).
33. D. Marculescu, R. Marculescu, N. H. Zamora, P. Stanley-Marbell, P. K. Khosla, S. Park, S. Jayaraman, S. Jung, C. Lauterbach, and W. Weber, *Proc. IEEE*, **91**, 1995 (2003).
34. L. V. Langenhove, in *Smart Textiles For Medicine And Healthcare: Materials, Systems And Applications* (Elsevier, Amsterdam) (2007).
35. M. Ghahremani Honarvar and M. Latifi, *J. Textile Institute*, **108**, 631 (2017).
36. T. Guinovart, M. Parrilla, G. A. Crespo, F. X. Rius, and F. J. Andrade, *Analyst*, **138**, 5208 (2013).
37. T. Guinovart, G. Valdés-Ramírez, J. R. Windmiller, F. J. Andrade, and J. Wang, *Electroanalysis*, **26**, 1345 (2014).
38. A. Abouraddy, M. Bayindir, G. Benoit, S. Hart, K. Kuriki, N. Orf, O. Shapira, F. Sorin, B. Temelkuran, and Y. Fink, *Nat. Mater.*, **6**, 336 (2007).
39. A. Hatamie, A. Khan, M. Golabi, A. P. Turner, V. Beni, W. C. Mak, A. Sadollahkhani, H. Alnoor, B. Zargar, S. Bano, O. Nur, and M. Willander, *Langmuir*, **31**, 10913 (2015).
40. M. Hamedi, R. Forchheimer, and O. Inganäs, *Nat. Mater.*, **6**, 357 (2007).
41. M. Gerard, A. Chaubey, and B. Malhotra, *Biosens. Bioelectron.*, **17**, 345 (2002).
42. D. Kumar and R. Sharma, *Eur. Polym. J.*, **34**, 1053 (1998).
43. J. Huang and R. B. Kaner, *Angewandte Chemie*, **116**, 5941 (2004).
44. K. Huang, M. Wan, Y. Long, Z. Chen, and Y. Wei, *Synth. Met.*, **155**, 495 (2005).
45. S. Park, M. Vosguerichian, and Z. Bao, *Nanoscale*, **5**, 1727 (2013).
46. L. Gonzalez-Macia, A. Morrin, M. R. Smyth, and A. J. Killard, *Analyst*, **135**, 845 (2010).
47. J. Weremczuk, G. Tarapata, and R. Jachowicz, *Procedia Engineering*, **47**, 1366 (2012).
48. E. Skrzetuska, M. Puchalski, and I. Krucińska, *Sensors*, **14**, 16816 (2014).
49. C. Xu, B. Xu, Y. Gu, Z. Xiong, J. Sun, and X. Zhao, *Energy & Environmental Science*, **6**, 1388 (2013).
50. Y. Liu, X. Dong, and P. Chen, *Chem. Soc. Rev.*, **41**, 2283 (2012).
51. L. Hu, M. Pasta, F. La Mantia, L. Cui, S. Jeong, H. D. Deshazer, J. W. Choi, S. M. Han, and Y. Cui, *Nano Lett.*, **10**, 708 (2010).

52. A. K. Yetisen, H. Qu, A. Manbachi, H. Butt, M. R. Dokmeci, J. P. Hinestroza, M. Skorobogatiy, A. Khademhosseini, and S. H. Yun, *ACS nano*, **10**, 3042 (2016).
53. R. Dastjerdi and M. Montazer, *Colloids Surf., B*, **79**, 5 (2010).
54. G. Nagaraju, Y. H. Ko, and J. S. Yu, *Cryst. Eng. Comm*, **16**, 11027 (2014).
55. B. Xu and Z. Cai, *Appl. Surf. Sci.*, **254**, 5899 (2008).
56. Z. L. Wang, *J. Phys. Condens. Matter*, **16**, R829 (2004).
57. V. Kaushik, J. Lee, J. Hong, S. Lee, S. Lee, J. Seo, C. Mahata, and T. Lee, *Nanomaterials*, **5**, 1493 (2015).
58. A. Pantelopoulos and N. G. Bourbakis, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, **40**, 1 (2010).
59. J. R. Windmiller and J. Wang, *Electroanalysis*, **25**, 29 (2013).
60. S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, *J. Neuroengineering Rehabil.*, **9**, 21 (2012).
61. P. Gründler, in *Chemical Sensors: An Introduction for Scientists and Engineers* (Springer Science & Business Media, Berlin) (2007).
62. X. Zhang, H. Ju, and J. Wang, in *Electrochemical Sensors, Biosensors And Their Biomedical Applications* (Academic, New York) (2011).
63. L. Jabbour, C. Gerbaldi, D. Chaussy, E. Zeno, S. Bodoardo, and D. Beneventi, *J. Mater. Chem.*, **20**, 7344 (2010).
64. L. Hu, H. Wu, F. La Mantia, Y. Yang, and Y. Cui, *ACS Nano*, **4**, 5843 (2010).
65. A. M. Gaikwad, D. A. Steingart, T. N. Ng, D. E. Schwartz, and G. L. Whiting, *Appl. Phys. Lett.*, **102**, 104 (2013).
66. S. Coyle, D. Morris, K.-T. Lau, D. Diamond, N. Taccini, D. Costanzo, P. Salvo, F. Di Francesco, M. G. Trivella, and J.-A. Porchet, *PervasiveHealth 2009, 3rd International Conf. on Pervasive Computing Technologies for Healthcare, 2009*, p. 1 (2009).
67. G. Baysal, S. Önder, İ. Göcek, L. Trabzon, H. Kızıl, F. N. Kök, and B. K. Kayaoğlu, *Sensors Actuators B*, **208**, 475 (2015).
68. V. F. Curto, C. Fay, S. Coyle, R. Byrne, C. O'Toole, C. Barry, S. Hughes, N. Moyna, D. Diamond, and F. Benito-Lopez, *Sensors Actuators B*, **171**, 1327 (2012).
69. B. Tang, Y. He, J. Liu, J. Zhang, J. Li, J. Zhou, Y. Ye, J. Wang, and X. Wang, *Dyes Pigm.*, **170**, 107643 (2019).
70. J. He, G. Xiao, X. Chen, Y. Qiao, D. Xu, and Z. Lu, *RSC Adv.*, **9**, 23957 (2019).
71. S. D. Bindesri, D. S. Alhatab, and C. L. Brosseau, *Analyst*, **143**, 4128 (2018).
72. N. Promphet, P. Rattanawaleedirojn, K. Siralertmukul, N. Soatthiyanon, P. Potiyaraj, C. Thanawattano, J. P. Hinestroza, and N. Rodthongkum, *Talanta*, **192**, 424 (2019).
73. R. E. Owyeung, M. J. Panzer, and S. R. Sonkusale, *Sci. Rep.*, **9**, 5607 (2019).
74. A. Modali, S. R. K. Vanjari, and D. Dendukuri, *Electroanalysis*, **28**, 1276 (2016).
75. K. Malzahn, J. R. Windmiller, G. Valdés-Ramírez, M. J. Schöning, and J. Wang, *Analyst*, **136**, 2912 (2011).
76. Y.-L. Yang, M.-C. Chuang, S.-L. Lou, and J. Wang, *Analyst*, **135**, 1230 (2010).
77. G. Lisak, T. Arnebrant, T. Ruzgas, and J. Bobacka, *Anal. Chim. Acta*, **877**, 71 (2015).
78. J. McCann and D. Bryson, in *Smart Clothes and Wearable Technology* (Elsevier, Amsterdam) (2009).
79. N. Coppedè, G. Tarabella, M. Villani, D. Calestani, S. Iannotta, and A. Zappettini, *J. Mater. Chem. B*, **2**, 5620 (2014).
80. J. Gonzalo-Ruiz, R. Mas, C. de Haro, E. Cabruja, R. Camero, M. A. Alonso-Lomillo, and F. J. Muñoz, *Biosens. Bioelectron.*, **24**, 1788 (2009).
81. M. Parrilla, J. Ferré, T. Guinovart, and F. J. Andrade, *Electroanalysis*, **28**, 1267 (2016).
82. R.-I. Stefan-van Staden, L. A. Gugoaşă, M. Badulescu, and C. C. Surdu-Bob, *RSC Adv.*, **5**, 45545 (2015).
83. G. E. Collins and L. Buckley, *Synth. Met.*, **78**, 93 (1996).
84. D. Kincal, A. Kumar, A. D. Child, and J. R. Reynolds, *Synth. Met.*, **92**, 53 (1998).
85. T. Kinkeldei, C. Zysset, N. Münzenrieder, and G. Tröster, *Sensors Actuators B*, **174**, 81 (2012).
86. D. S. Kinnamon, S. Krishnan, S. Brosler, E. Sun, and S. Prasad, *J. Electrochem. Soc.*, **165**, B3084 (2018).
87. X. Wang, Y. Gu, Z. Xiong, Z. Cui, and T. Zhang, *Adv. Mater.*, **26**, 1336 (2014).
88. B. Schazmann, D. Morris, C. Slater, S. Beirne, C. Fay, R. Reuveny, N. Moyna, and D. Diamond, *Anal. Methods*, **2**, 342 (2010).
89. A. J. Bandodkar and J. Wang, *Trends Biotechnol.*, **32**, 363 (2014).
90. T. Hughes-Riley, P. Lugoda, T. Dias, C. Trabi, and R. Morris, *Sensors*, **17**, 1804 (2017).
91. S. K. Bahadir, V. Koncar, and F. Kalaoglu, *Sens. Actuators, A*, **179**, 297 (2012).
92. R. Paradiso, G. Loriga, and N. Taccini, *IEEE Trans. Inf. Technol. Biomed.*, **9**, 337 (2005).
93. H. Mattila, in *Intelligent Textiles and Clothing* (Woodhead Publishing, Cambridge, England) (2006).
94. S. Bielska, M. Sibinski, and A. Lukasik, *Materials Science and Engineering: B*, **165**, 50 (2009).
95. C. Yan, J. Wang, and P. S. Lee, *ACS nano*, **9**, 2130 (2015).
96. Y. Wang, H. Guo, J.-J. Chen, E. Sowade, Y. Wang, K. Liang, K. Marcus, R. R. Baumann, and Z.-S. Feng, *ACS Applied Materials Interfaces*, **8**, 26112 (2016).
97. J. Huang, Y. Hao, H. Lin, D. Zhang, J. Song, and D. Zhou, *Mater. Sci. Eng. B*, **99**, 523 (2003).
98. Y. Yang, Z.-H. Lin, T. Hou, F. Zhang, and Z. L. Wang, *Nano Res.*, **1**, 888–895 (2012).
99. P. Shyamkumar, P. Rai, S. Oh, M. Ramasamy, R. E. Harbaugh, and V. Varadan, *Electronics*, **3**, 504 (2014).
100. J. Yoo, L. Yan, S. Lee, H. Kim, and H.-J. Yoo, *IEEE Trans. Inf. Technol. Biomed.*, **13**, 897 (2009).
101. Y. Li, Y. Li, M. Su, W. Li, Y. Li, H. Li, X. Qian, X. Zhang, F. Li, and Y. Song, *Advanced Electronic Materials*, **3**, 1700253 (2017).
102. Y. Li, Y. A. Samad, and K. Liao, *Journal of Materials Chemistry A*, **3**, 2181 (2015).
103. G. Sainarayanan, R. Nagarajan, and S. Yaacob, *Appl. Soft Comput.*, **7**, 257 (2007).
104. J. Lee, H. Kwon, J. Seo, S. Shin, J. H. Koo, C. Pang, S. Son, J. H. Kim, Y. H. Jang, and D. E. Kim, *Adv. Mater.*, **27**, 2433 (2015).
105. S. Takamatsu, T. Kobayashi, N. Shibayama, K. Miyake, and T. Itoh, *Sens. Actuators, A*, **184**, 57 (2012).
106. K. Cherenack, C. Zysset, T. Kinkeldei, N. Münzenrieder, and G. Tröster, *Adv. Mater.*, **22**, 5071 (2010).
107. Y. Wang, J. Zheng, G. Ren, P. Zhang, and C. Xu, *Smart Mater. Struct.*, **20**, 045009 (2011).
108. M. Rothmaier, M. P. Luong, and F. Clemens, *Sensors*, **8**, 4318 (2008).
109. J. Meyer, P. Lukowicz, and G. Troster, *10th IEEE International Symposium on Wearable Computers*, p. 69 (2006).
110. J. Meyer, *Textile Pressure Sensor: Design, Error Modeling and Evaluation*, ETH Zurich (2008).
111. J. Karvonen and T. Vuorimaa, *Sports Med.*, **5**, 303 (1988).
112. R. M. Laukkanen and P. K. Virtanen, *J. Sports Sci.*, **16**, 3 (1998).
113. H. J. Lee, S. H. Hwang, H. N. Yoon, W. K. Lee, and K. S. Park, *Sensors*, **15**(5), 11295-11311 11295 (2015).
114. J. Coosemans, B. Hermans, and R. Puers, *Sensors Actuators A*, **130**, 48 (2006).
115. P. Guay, S. Gorgutsa, S. LaRochelle, and Y. Messaddeq, *Sensors*, **17**, 1050 (2017).
116. O. Atalay, W. R. Kennon, and E. Demirok, *IEEE Sensors J.*, **15**, 110 (2015).
117. C.-T. Huang, C.-L. Shen, C.-F. Tang, and S.-H. Chang, *Sensors Actuators A*, **141**, 396 (2008).
118. T. Hoffmann, B. Eilebrecht, and S. Leonhardt, *IEEE Sensors J.*, **11**, 1112 (2011).
119. S. D. Min, Y. Yun, and H. Shin, *IEEE Sensors J.*, **14**, 3245 (2014).

Utilizing Temporal Psycholinguistic Cues for Suicidal Intent Estimation

Puneet Mathur*[1], Ramit Sawhney*[2], Shivang Chopra[3], and Rajiv Ratn Shah[4]

[1] University of Maryland, College Park, USA
[2] Netaji Subhas Institute of Technology, Delhi, India
[3] Delhi Technological University, Delhi
[4] MIDAS Labs, IIIT Delhi, India

Abstract. Temporal psycholinguistics can play a crucial role in studying expressions of suicidal intent on social media. Current methods are limited in their approach in leveraging contextual psychological cues from online user communities. This work embarks in a novel direction to explore historical activities of users and homophily networks formed between Twitter users for extracting suicidality trends. Empirical evidence proves the advantages of incorporating historical user profiling and temporal graph convolutional modeling for automated detection of suicidal connotations on Twitter.

1 Introduction

Suicidal ideation detection is a well studied problem in social media analysis. Various works have tried to identify linguistic patterns correlated with suicidality intent. Despite the sustained efforts from the community, most approaches ignore the psychological relevance of temporal characteristics of suicidal behaviour. Moreover, there has been limited explorations in the space of homophily networks to identify collusive depressive users. We hypothesize that the contextual information embedded in social media engagement and historical activities of users can lead to substantial improvements in automated identification of suicidal ideation. We look beyond linguistic cues into temporal signals throughout this work, with the help of a publicly available dataset given by [14] of 34,306 tweets on suicidality detection.

2 Related Work

2.1 Challenges on Social Media

The growth of social media websites hosts a number of challenges such as cyberbullying, suicide pacts, and radicalism that motivate suicidal behavior and impact the mental health of the users [10]. The associativity of suicide-related

* equal contribution

verbalizations on social media websites has been found to be strongly related to potential suicidal attempts. Prior studies show how suicidal intent declarations were significantly more assortative than chance, at times connected till 6 degrees of separation [5]. A patient's social media profile can help medical experts gain perspective into their mental health status and identify those at critical risk for suicide attempts [15]. The potential of technological interventions for suicidal risk assessment and mitigation needs to be explored in detail.

2.2 Text-based Approaches

Various works have been recently proposed with an objective of automating the detection of social media posts expressing suicide ideation using textual information [17,7,3]. [4] performed a semi-automated content-based analysis on a small number of tweets related to depression in order to derive certain qualitative insights into the behavior of users displaying suicidal behavior. Self-disclosure helps to facilitate psychological well being in individuals with mental illness [2]. Textual descriptions of social media disclosures have been extensively studied in the past [7]. [19] explores deep learning based supervised classifiers for suicidal ideation detection.

2.3 Psycho-linguistic Analysis

[13] used social graph based features and gained considerable improvement in the task of abuse detection. [16] performed a psycho-linguistic analysis of online users for a similar task. [1] tries to link users' psychological features such as personality traits including personalities, sentiment and emotion for cyberbulling and trolling. The contributions that we make in this work are different from previous efforts as there has been hardly any attempt to take a combined multi-faceted approach for solving the task of suicidal ideation in Twitter.

2.4 Signals from Temporal Data

Temporal graphs can capture the relationships in data with time so as to model new events and comparison to related entities and historical states [18]. [9] detected groups based on interesting features of the time-evolving networks. It studied several clustering frameworks for time-evolving networks for detecting group structure. [6] performed temporal sentiment analysis for early detection of cyberbulling and suicide ideation of a user through graph-based data mining approaches.

3 Methodology

The proposed methodology looks beyond text classifiers and leverages tweeting history of users as well as their social network communication patterns. User-based features were extracted from the historical tweeting activity and inter-user interactions was modeled as a social graph. The methodology is two-fold consisting of historical signal modeling and temporal graph convolutional modeling.

3.1 Classification Network

In order to learn from the textual information available in the raw tweets, we trained a **BLSTM + Attention** network [20]. We train a BLSTM model with 100 LSTM units, dropout rate of 0.25 and a recurrent dropout rate of 0.2. The attention layer was followed by another dropout layer of 0.2. This was followed by two dense layers having 256 units and 2 units, respectively.

3.2 Temporal Modeling of Suicidal Tendency

Motivation: The idea of temporal modeling of suicidal tendencies is inspired by [11] with additions. According to [11], a representation for the historical activity can be formulated as a temporal weighting scheme ϕ_i which is a sum of two independent time varying functions of suicidality - ideation build-up $\lambda_i(t)$ and sinusoidal episodes $\mu_i(t)$. Extrapolating from this, we add a third independent time-varying function - white Gaussian noise $z_i(t)$. Let Δt_i be the time offset from the original tweet and the temporal representation function z be given by Equation 1.

$$z(u, H) = \sum_{h_i \in H} \phi_i(\Delta t) f(h_i) \tag{1}$$

Suicidal Ideation Build-up: Each user's historical tweets can be modeled as an exponential function in time given by equation 2 where α and β are hyper parameters tuned over training data.

$$\lambda_i(\Delta t) = \alpha e^{\beta \Delta t_i} \tag{2}$$

Suicidal Episodes: Phased changes in suicidal intent are mathematically represented by equation 3. As per [12], the hyper parameters for the same are given by Table 1.

$$\mu_i(\Delta t) = \sum_{1}^{Q} (a_q cos(\frac{2\pi q \Delta t_i}{U}) + b_q sin(\frac{2\pi q \Delta t_i}{U})) \tag{3}$$

Hyperparameters	**Value**
Q	3
U	{1,2,3,4,5,6,7}
a_q , b_q	$\approx \eta(0, \sigma^2)$

Table 1: Hyper parameters for Equation 3 [11]

Temporal surprise: Similar to any channel medium, social media platforms are prone to noise that adds randomness to the temporal suicidal patterns. The

white Gaussian noise is modeled as being derived from a normal distribution with the expectation value of the noise term $\zeta_(t)$ equal to 1.

$$\phi_i(\Delta t) = \lambda_i(\Delta(t) + \mu_i(\Delta t) + \zeta_i(\Delta t) \tag{4}$$

For each of the tweet samples, the historical activity representation was an input to logistic regression model to learn temporal embeddings from these features which was used as an input to the final model.

3.3 Graph Convolutional Networks for User Profiling

Learning user representations can be significantly enriched by leveraging information derived from the inter-user interactions in social media channels. For this purpose, Graph Convolutional Networks (GCN) [8] can be effectively utilized that are capable of modeling social interactions in the form of features of nodes in the graph and allow contextual learning of information with respect to a node's neighbourhood.

Temporal GCN: We tried to incorporate the historical views into the extended graph by constructing time weighted TF-IDF vectors of the historical tweets. The author nodes were modified to consist of temporal weighting of TF-IDF representation of tweets. Let the TF-IDF vector f_k^t of tweet at timestamp t for k^{th} author be defined by Equation 5, where C_k is the global noise parameter, $\hbar$ controls the margin of influence of a user on its neighbours social activity and ω is the rate of decay of the suicidal sentiment. The external parameters C_k, $\hbar_k$ and ω_k are learnt from the training portion of the dataset in an unsupervised fashion.

$$f^t = \hbar_k \exp^{\omega_k \Delta t} \tag{5}$$

4 Experiments

4.1 Data Description and Setup

To gauge the effectiveness of our proposed approach, we use the dataset from SNAP-BATNET[14] which consists of $34,306$ tweets with $3,984$ of them suicidal ideations. For each of these users, the tweet timelines were also collected to create the set of historical tweets. 10-fold stratified cross-validation was employed to evaluate models on each of the 10 train-val splits. The hyper parameters for the temporal weighted combination were tuned using a grid search over the grid $\alpha = \{0.1, 0.5, 1.0\}$, $\beta = \{0, 0.01, 0.1, 1\}$, $U = \{1, 2..., 7\}$ yielding $\alpha = 0.5, \beta = 1, U = 7$. t_0 was assigned to time series points with values equal to $argmax(|\mu|)_i$.

Model	F1	P	R
SNAP-BATNET [14]	92.60	72.20	93.52
BiLSTM + Attention (Text)	91.26	70.02	91.23
Text + Temporal Modeling	92.75*	91.98*	93.70*
Temporal GCN	**93.89***	88.73	**94.54***

Table 2: Performance Analysis

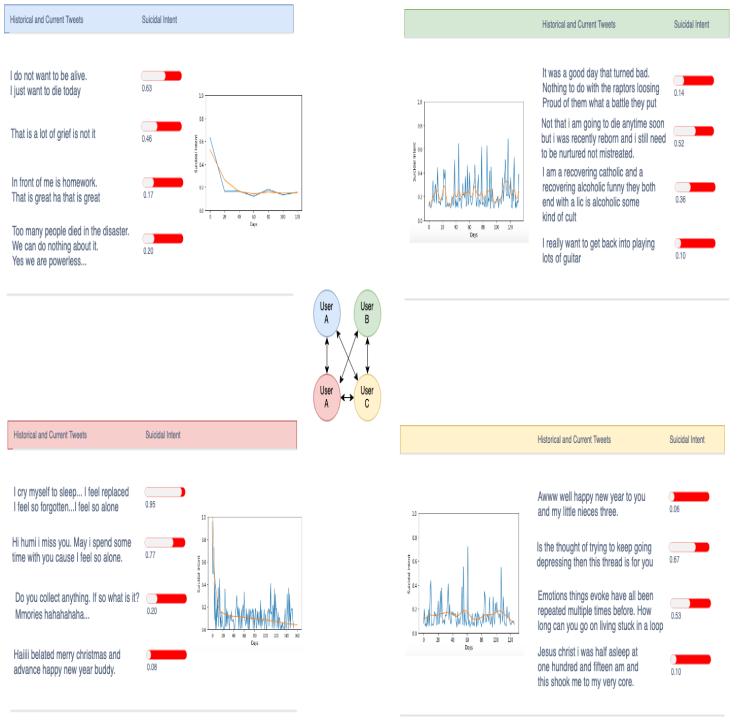


Fig. 1: Analysis of historical behaviour of users in a community over time

5 Results and Ablation Analysis

The ablation study of experimented features presented in Table 2 highlights the significance of temporal features extracted from social media in suicide ideation risk assessment. Temporal GCN provides a substantial gain over text in prediction confidence due to the user interactions. Additionally, it is interesting to observe the ability of the GCN model to better represent historical suicidal signals in comparison to naive historical and textual features to a sufficient degree. Empirically, temporal features help suppress false positives induced by text classifiers that try to overfit on the presence of anecdotal suicidal phrases such as "*kill me...hahaha !!*" that may be considered as noise in non-suicidal text. The most optimal weights for temporal signal modeling **Text + Builtup + Episodic + Surprise** were derived to be 0.52, 0.04, 0.04 and 0.32 through cross-validation experiments.

Figure 1 elucidates the impact of including psychological contextual cues on a small sample of connected users from the test dataset. It is evident from the historic trends of Users B and C that they follow a nearly episodic nature with scattered surprises. Analysing the trend plots for Users A and D reveals an inverse build-up thereby demonstrating that there can be either a positive or negative build-up in the suicidal intent of users. All these aspects when captured by our model has led to a statistically significant increase in the model's performance.

6 Conclusion

In spite of high importance of suicidal ideation identification on social media, little research has focused on looking beyond linguistic patterns. Through our work, we demonstrate that user interactions and past user behaviour are strong indicators of a potentially concerning mental state of online users. In this study, employing both qualitative and quantitative methods, we address this gap by investigating the impact of augmenting text based suicidal ideation detection models with contextual cues based on historical tweeting behavior and social media engagement.

References

1. Balakrishnan, V., Khan, S., Arabnia, H.R.: Improving cyberbullying detection using twitter users' psychological features and machine learning. Computers & Security p. 101710 (2020)
2. Balani, S., De Choudhury, M.: Detecting and characterizing mental health related self-disclosure in social media. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems. pp. 1373–1378. ACM (2015)
3. Benton, A., Mitchell, M., Hovy, D.: Multi-task learning for mental health using social media text. arXiv preprint arXiv:1712.03538 (2017)

4. Cavazos-Rehg, P.A., Krauss, M.J., Sowles, S., Connolly, S., Rosas, C., Bharadwaj, M., Bierut, L.J.: A content analysis of depression-related tweets. Computers in human behavior **54**, 351–357 (2016)
5. Cero, I., Witte, T.K.: Assortativity of suicide-related posting on social media. American Psychologist (2019)
6. Chatterjee, A., Das, A.: Temporal sentiment analysis of the data from social media to early detection of cyberbullicide ideation of a victim by using graph-based approach and data mining tools. In: Intelligence Enabled Research, pp. 107–112. Springer (2020)
7. De Choudhury, M., Gamon, M., Counts, S., Horvitz, E.: Predicting depression via social media. In: Seventh international AAAI conference on weblogs and social media (2013)
8. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
9. Lee, K.H., Xue, L., Hunter, D.R.: Model-based clustering of time-evolving networks through temporal exponential-family random graph models. Journal of Multivariate Analysis **175**, 104540 (2020)
10. Lopez-Castroman, J., Moulahi, B., Azé, J., Bringay, S., Deninotti, J., Guillaume, S., Baca-Garcia, E.: Mining social networks to improve suicide prevention: A scoping review. Journal of neuroscience research (2019)
11. Mathur, P., Sawhney, R., Shah, R.R.: Suicide risk assessment via temporal psycholinguistic modeling (student abstract). In: Proceedings of the 34th AAAI Conference on Artificial Intelligence 2020. AAAI (2020)
12. Mathur, P., Shah, R., Sawhney, R., Mahata, D.: Detecting offensive tweets in hindi-english code-switched language. In: Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media. pp. 18–26 (2018)
13. Mishra, P., Del Tredici, M., Yannakoudakis, H., Shutova, E.: Author profiling for abuse detection. In: Proceedings of the 27th International Conference on Computational Linguistics. pp. 1088–1098 (2018)
14. Mishra, R., S.P.S.R.M.D.M.P., Shah, R.: Snap-batnet: Cascading author profiling and social network graphs for suicide ideation detection on social media. Proceedings of the 2019 NAACL Student Research Workshop (pp. 147-156). (2019)
15. Pourmand, A., Roberson, J., Caggiula, A., Monsalve, N., Rahimi, M., Torres-Llenza, V.: Social media and suicide: a review of technology-based epidemiology and risk assessment. Telemedicine and e-Health **25**(10), 880–888 (2019)
16. Qian, J., ElSherief, M., Belding, E.M., Wang, W.Y.: Leveraging intra-user and inter-user representation learning for automated hate speech detection. arXiv preprint arXiv:1804.03124 (2018)
17. Sawhney, R., Manchanda, P., Mathur, P., Shah, R., Singh, R.: Exploring and learning suicidal ideation connotations on social media with deep learning. In: Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. pp. 167–175 (2018)
18. Steer, B., Cuadrado, F., Clegg, R.: Raphtory: Streaming analysis of distributed temporal graphs. Future Generation Computer Systems **102**, 453–464 (2020)
19. Tadesse, M.M., Lin, H., Xu, B., Yang, L.: Detection of suicide ideation in social media forums using deep learning. Algorithms **13**(1), 7 (2020)
20. Zhou, P., Qi, Z., Zheng, S., Xu, J., Bao, H., Xu, B.: Text classification improved by integrating bidirectional lstm with two-dimensional max pooling. arXiv preprint arXiv:1611.06639 (2016)