SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(73rd Edition)

JANUARY, 2018

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 **Seventy Third** 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, 2018**. 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 Gene Family-led Meta-Analysis of Drug-Target Interactions, Qiong Cheng, ***3.Saurabh Mehta*** and Stephan Schiirer, Chemistry, DTU

2. A mathematical and numerical study of a SIR epidemic model with time delay, nonlinear incidence and treatment rates, ***6.Kanica Goel*** and ***3.Nilam***, Mathematics

3. Adaptive Technique with Cross Correlation for Lowering Signal-to-Noise Ratio Wall in Sensor Networks, ***3.Rajiv Kapoor***, Rashmi Gupta, Le Hoang Son, Sudan Jha and Raghvendra Kumar, Electronics & Communication, DTU

4. Applicability of Machine Learning in the Measurement of Emotional Intelligence, ***6.Manish Sharma***, ***3.Shikha N. Khera*** and Pritam B. Sharma, DSM, DTU

5. Approximation by genuine Gupta–Srivastava operators, ***6.Ram Pratap*** and ***3.Naokant Deo***, Applied Mathematics, DTU

6. ARTIFICIAL INTELLIGENCE: DEVELOPING SOFTWARE-BASED APPLICATION TO LEVERAGE GENETIC ALGORITHM TO ENHANCE ITS APPLICABILITY IN BUSINESS AND MEDICAL AREAS, ***8.Ruchika Goel***, Computer, DTU

7. Automated System for Prediction of Skin Disease using Image Processing and Machine Learning, ***6.Lakshay Bajaj***, ***3.Himanshu Kumar*** and ***3.Yasha Hasija***, Biotechnology, DTU

8. Adaptive Technique with Cross Correlation for Lowering Signal-to-Noise Ratio Wall in Sensor Networks, ***3.Rajiv Kapoor***, Rashmi Gupta, Le Hoang Son, Sudan Jha and Raghvendra Kumar, Electronics & Communication, DTU

9. Breast cancer incidence, mortality and mortality-to-incidence ratio (MIR) are associated with human development, 1990–2016: evidence from Global Burden of Disease Study 2016, ***3.Rajesh Sharma***, DTU

10. Comparison of half effect absorption-compression cascaded refrigeration system using thermodynamic (energy-exergy) analysis, ***3.R.S. Mishra***, DTU

11. Cyberbullying detection on social multimedia using soft computing techniques: a meta-analysis, ***3.Akshi Kumar*** & ***6.Nitin Sachdeva***, Computer Science, DTU

12. Design and Analysis of Complimentary Split Ring Resonator Backed Microstrip Transmission Line Using Equivalent Circuit Model, ***6.P. Garga*** and ***3.P. Jaina***, ECE, DTU

13. Detection of Ovarian Cancer using Organic Light Emitting Diodes, Shubham Negi, ***3.Poornima Mittal***, Brijesh Kumar and Anamika Bhatia, ECE, DTU

14. ELECTROCHEMICAL PAPER BASED CANCER BIOSENSOR USING IRON OXIDE NANOPARTICLES DECORATED PEDOT:PSS, ***6.Saurabh Kumar***, ***8.Mohammad Umar***, ***6.Anas Saifi***, ***6Suveen Kumar***, ***6.Shine Augustine***, ***6.Saurabh Srivastava***, ***3.Bansi D Malhotra,*** Biotechnology, DTU

15. Embedded CNN based vehicle classification and counting in non-laned road traffic, Mayank Singh Chauhan, Arshdeep Singh, ***8.Mansi Khemka***, Arneish Prateek and Rijurekha Sen, Computer Science, DTU

16. Empowerment of Women in India, ***6.Neha Tomar*** and ***3.Seema Singh***, Humanities, DTU

17. EXPERIENCE OF WOMEN IN ENGINEERING EDUCATION: LITERATURE REVIEW, ***6.Addissie Melak*** and ***3.Seema Singh***, Humanity, DTU

18. Improving the consistency of incomplete hesitant multiplicative preference relation, ***6.Mamata Sahu*** and ***3.Anjana Gupta***, Applied Mathematics, DTU

19. Lossless grounded admittance simulator using OTRA, Bal Chand Nagar and ***6.Sajal K. Paul***, Electronics and Communication, DTU

20. Numerical Investigation of Autoignition of Singlecomponent and Multicomponent surrogate of Diesel fuel, ***6.Anirudh Singh*** and ***3.Amit Pal***, Mechanical, DTU

21. Object Detection in Foggy Conditions by Fusion of Saliency Map and YOLO, ***8.Sarthak Katyal***, ***8.Sanjay Kumar***, ***8.Ronak Sakhuja*** and ***8.Samarth Gupta***, CSE, DTU

22. Optimization of Material Removal Rate in Wire-EDM using Genetic Algorithms, G Daniel, ***3.Dr. Girish kumar*** and Dr Ali-Ul-Hasan-Rizvi, Mechanical, DTU

23. Optimum Preventive Maintenance Policy for a Mechanical System Using Semimarkov Method and Golden Section Technique, ***3.G. Kumar*** and ***6.J. P. Varghese***, Mechanical, DTU

24. Particle Swarm Optimisation of Hole Quality Characteristics in Laser Trepan Drilling of Inconel-718, ***7.1.Kedari Lal Dhaker*** and Arun Kumar Pandey, Production, DTU

25. Performance Analysis of Different Bootstrap Inverter Configurations using Dual Gate OTFT, Neha Mishra, ***3.Poornima Mittal*** and Brijesh Kumar, ECE, DTU

26. Performance improvement of vapour compression refrigeration systems, ***3.R.S. Mishra***, Mechanical & Production, DTU

27. Physical vulnerability assessment of buildings exposed to landslides in India, Aditi Singh, D. P. Kanungo and ***3.Shilpa Pal***, Civil, DTU

28. Principal Component Analysis Of TF-IDF In Click Through Rate Prediction, ***8.1.Ankita Pal***, Mathematics, DTU

29. Semi Supervised Graph Based Keyword Extraction Using Lexical Chains and Centrality Measures, ***8.Ayush Aggarwal***, ***8.1.Chhavi Sharma***, ***3.Minni Jain*** and Amita Jain, Computer, DTU

30. Solar Potential Analysis of Rooftops Using Satellite Imagery, ***6.Akash Kumar*** and ***3.S. Indu***, Electronics and Communication, DTU

31. Step Up Gain Converter with fast MPPT control under moving partial shading for train rooftop PV-DC-μG, ***6.Shirish Raizada*** and ***3.Vishal Verma***, Electrical, DTU

32. Study of Element Transfer for Silica Based Flux Constituents by Taguchi Analysis in Submerged Arc Welding, ***7.1.Aditya Kumar***, Mechanical, DTU

33. Technology forecasting (TF) and technology assessment (TA) methodologies: a conceptual review, Abid Haleem, Bisma Mannan, Sunil Luthra, ***7.1.Sanjay Kumar*** and Sonal Khurana, Mechanical and Production Engineering, DTU

34. The effect of odd-even driving scheme on PM2.5 and PM1.0 emission, ***3.Rajeev Kumar Mishra***, ***6.Abhinav Pandey***, Govind Pandey and ***6.Amrit Kumar***, Environmental, DTU

35. The Role of Big Data Predictive Analytics and Radio Frequency Identification in the Pharmaceutical Industry, MUHAMMAD NOMAN SHAFIQUE, MUHAMMAD MAHBOOB KHURSHID, HAJI RAHMAN, ASHISH KHANNA, AND ***7.1.DEEPAK GUPTA***, Computer Science, DTU

36. UWB Metamaterial-Loaded Antenna for C-Band Applications, Parul Dawar, ***3.N. S. Raghava*** and ***3.Asok De***, Electronics & Communication, DTU

37. Vibration analysis for detecting failure mode and crack location in first stage gas turbine blade, ***3.S. Rani***, ***3.A. K. Agrawal*** and ***3.V. Rastogi***, Mechanical, Production, Industrial and Automobile, DTU

38. Wavefunction Engineering of Type-I/Type-II Excitons of CdSe/CdS Core-Shell Quantum Dots, ***8.Yashaswi Nandan*** & ***3.Mohan Singh Mehata***, Applied Physics, DTU

39. Writer Independent Oine Signature Recognition Using Ensemble Learning, Sourya Dipta Das, ***8.Himanshu Ladia***, ***8.Vaibhav Kumar*** and ***8.Shivansh Mishra***, Mathematics and Computer Science, 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 Gene Family-led Meta-Analysis of Drug-Target Interactions

Qiong Cheng
Department of Computer Science
University of North Carolina at Charlotte
NC, USA
qcheng1@uncc.edu

Saurabh Mehta
Department of Applied Chemistry
Delhi Technological University Delhi,
India
saurabh.dtu@gmail.com

Stephan Schürer
Department of Molecular and Cellular Pharmacology
University of Miami
FL, USA
sschuerer@med.miami.edu

***Abstract*— The shift from "magic bullets" to "magic shotguns" has prospered the pharmaceutical industry, where a belief is that a drug that "hits" multiple sensitive nodes belonging to a network of interacting targets offers the potential for higher efficacy and may limit drawbacks arising from the use of a single-target drug or a combination of multiple drugs. More high-quality drug-target interactions, profiled by the "magic shotguns" design paradigm, have emerged from diverse labs. However, the rapid production at scale and in variety challenges the efficacy of data usage. A creative and systematical approach is in demand to comprehensively analyze and integrate drug-target interactions for better prediction.**

In the project, taking the diversity into consideration, we carried out a gene family-led, meta-analysis, investigated diverse properties that drug-target pharmacological promiscuity relies on, and examined the consistency or integrity of drug-target data. The novel approach can facilitate the identification of cohesive multi-target combinations and revealed the interconnected properties of determining and rationalizing the promiscuity of profiled drugs. The approach can be further expanded to coordinate other experimental drug-target data and set a stage for the analysis of the mechanism of action of biological therapies.

Keywords*— *polypharmocology, ontology, drug-target interactions, information content, meta-analysis

I. INTRODUCTION

A mainstream drug discovery process is an assembly pipeline starting with target identification, following with lead identification and optimization, and completing with preclinical and clinical assessment [62]. In the process, target selection is a key factor for a success drug; drug efficacy and safety are the main consideration.

Numerous target-based drug profiling datasets have been generated through the efforts of a panel of protein targets against drug compound libraries. They are publically available via different databases such as NIMH Psychoactive Drug Screening Program (PDSP) data [1, 2], Kinome Knowledgebase (KKB) [19], KINOMEscan, and ChEMBL [22].

It is challenging for a data coordinator center to systematically collect and manage the data as well as execute data governance with diversity in consideration, especially when a drug-target experimental data may be biasedly towards a subset of proteins being selected to screen against a chemical library [68, 44]. Unpredicted drug toxicity happens when candidate compound binds to and influences not only the desired target but also other unintended proteins, which cause side effects [63-66]. Identifying and prioritizing on/off targets is challenging and plays a critical role in drug discovery.

We propose a knowledge-driven approach and software infrastructure to facilitate data governance. In the paper, we focus on cross validation of the efficacy of Drug Target Ontology (DTO) and examine the consistency or integrity of drug-target interaction data at diverse levels for better data governance.

First, we used the druggable genome described in the IDG project [51, 59-62, 64] as the reference to examine the number of selected targets in a drug target data set. We further visualized its coverage in the classification tree of drug targets presented in the DTO [59], which underlying gene family classification structure provides a global functional view of target classes. We also employed the statistical method to measure whether compounds interacting with targets within the same protein category of DTO are more similar to those across different categories. We expanded the knowledge driven approach to explore drug promiscuity and calculate the relatedness in terms of the chemical structure similarity of related drug sets.

II. MATERIALS, METHODS, AND RESULTS

A. Data

Our interest in the integrative analysis has been inspired by the early successful replacement of "magic bullets" with "magic shortguns" in treating psychiatric diseases [7,9]. Among various databases, PDSP (NIMH Psychoactive Drug Screening Program) database is a database of binding affinities (Ki) of drugs and chemical compounds for receptors, neurotransmitter transporters, ion channels, and enzymes, which is profiled in diverse species [1-2]. It includes ~55.4k ligand-receptor combinations in the version (kidb110121), where a ligand is defined as active against a protein target if and only if the binding affinity ki<10□M. After adding organisms from UniProt and also adding Human orthologs, we obtained 3934 unique ligands with SMILES (MW > 50) modulating 428 unique UniProt IDs.

ChEMBL is a database covering a wide range of curated and annotated data, manually extracted from the primary published literature and databases. The database collects the rich information of binding, functional and ADMET (absorption, distribution, metabolism, excretion and toxicity) for a large number of drug-like bioactive compounds [22] and covers targets ranging from single proteins, to protein complexes, then tissues and finally whole organism in vivo data. It includes ~5.4 million bioactivity measurements for more than 1 million compounds and ~5200 protein targets. We extracted all binding bioassay measurements for the LINCS compounds [23], which show the confidence level greater than or equal to 7. There are three binding types, such as Ki, Kd, or IC50. We aggregated all those measurements with the negative logarithm value greater than or equal to 5, along with mapping the target ChEMBL ID to UniprotID and then gene symbol (homo sapiens).

B. *Visualization of studied drug targets under the reference of drug target classification showing that studied drug targets are highly biased*

In order to effectively analyze and illustrate the coverage of small molecule that bind to the human druggable genome, we employ Drug Target Ontology (DTO) [59] as the reference of druggable target knowledgebase (version DTO_merged_v09_b3101), where a relatively complete collection of disease-linked human proteins and their belonging classification schema are enclosed. The classification of druggable proteins has been represented as in a tree-like hierarchical structure (also called subsumption tree), where leaf nodes represent as protein targets, non-leaf nodes as classes (also called categories), and edges adjacent to two nodes as the "is-a" subsumption relationship [59]. For example, ADRA1A is an alpha adrenoceptor (adrenergic receptor), which is a member of the GPCR superfamily that recognizes the endogenous ligand adrenaline and noradrenaline and which are amines and thus is also classified as aminergic receptor [58]. In the subsumption "is-a" hierarchy, upper-level target class subsumes its lower-level target classes.

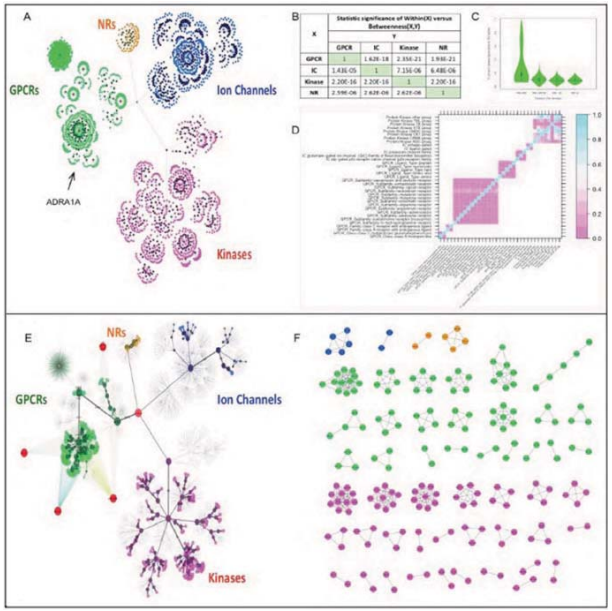


Figure 1. DTO drug target classifications facilitated the integrative analysis of PDSP binding bioactivity data. A. Visualization of DTO drug targets and its classification tree. **B.** Target-driven active PDSP ligands distinguish IDG families. We calculated the similarity of active compounds driven by targets in the same IDG superfamily and also in different superfamilies. We further measured the statistical significance of population difference by t-test. It turns out the active compounds against targets in the same protein family are more similar than in different families. **C.** Distribution of the similarity scores of pair-wise active compound profiled against targets in the same group against in different groups as an example. **D.** Statistical significance of the specificity of DTO classification. We measured the statistical significance for families or subfamilies of interest by using the same method as B. **E.** The coverage of active protein targets selected by PDSP compound library on DTO (Diverse GPCRs' annotations have been stressed, for example, ligand-binding receptor, amino acid derived receptor, amine associated receptor, peptide-binding receptor). **F.** Concordant target-to-target relatedness network, taking account of both related ligand set similarity and DTO target classification hierarchical structure similarity. Red nodes represent kinases, green GPCRs, blue ion channels, and orange nuclear receptors.

We visualized the DTO subsumption tree by the force-directed graph-drawing algorithm plugged in Cytoscape [31, 67] (Figure 1A), where we illuminate different druggable protein families by colors. There are four colors, and each color stands for each superfamily (also called an IDG family, for example, G protein-coupled receptors (GPCRs), kinases, ion channels (ICs), and nuclear receptors (NRs)). Protein nodes are colored in light and protein class nodes in dark. In the reference graph, each node has the uniform size. In addition to the "is-a" relationship, we also included other relationships in order to characterize both chemical structural, and pattern-based or ligand-bound annotation of protein targets. For example, ADRA1A binds an amine endogenous ligand;GRM5 is bound to an amino-acid ligand.

We downloaded the PDSP data, curated, performed a cross-reference to Uniprot IDs, and extracted the active drug-target combinations with a cutoff of Ki < 10uM, and then aggregated active compounds-hitting targets in each category. In order to demonstrate the distribution of those compounds-hit targets in druggable proteome, we use the above-mentioned reference graph as the background with opacity = 0.5 and have highlighted those nodes of compound-interacting targets in PDSP and its belonging classes (see Figure 1E). For each active target or category, we showed the number of active compounds on the edge adjacent to its node. We emphasized the hierarchical structure and active compounds in different target classes by designing the weighted scoring scheme to scale the node size and edge width, which takes account of the size of the category, the level of the category, and the logarithm of the number of active compounds hitting targets.

The knowledge-driven approach to visualize the experimental coverage, as shown in Figure 1A, and 1E, illustrates areas with very limited experimental exploration, which may provide opportunities for new drug development efforts. For example, all 424 olfactory receptor proteins are not profiled in both screening data (see Discussion).

C. Ligand drug target specificity and selection

We also analyzed the number of protein targets profiled against each PDSP compound. Out of 16,969 aggregated human small molecule-target combinations, we observed that 71.2% drugs have been tested against more than one target and 20.2% compounds have been screened by different laboratories, designating them as "compounds of interest". We note that a large number of compounds have been tested biasedly towards a subset of druggable proteome. The observed selectivity bias may reasonably arise from the technical expertise or prior knowledge about the druggable proteome.

D. Relationship between target class and molecular properties

We calculated a set of Tanimoto (Tc) scores as a measure of molecular properties for all pairwise active compounds in the drug-target data in order to investigate the relationship between target class and molecular properties of related active compounds. Tanimoto (Tc) scores are measured under Atom type Extended Connectivity Fingerprints with a diameter of 4 (ECFP4). Distinct differences are observed when comparing the distribution of Tc scores between sets of compounds against same IDG subfamilies with those against different IDG subfamilies (see Figure 1B, 1C). Then we measured the statistical significance of the distribution difference, namely, the occurrence possibility of the mean Tc score of active compounds of intra-target class greater than inter-target classes by chance. Figure 1C takes nuclear receptors as an example, showing the distribution of intra-class and inter-class, where the interquartile range, median, min, max, and density are shown. Through performing the unpaired two-sample one-tailed t-test, we examined that the population mean of the Tc scores calculated within target class is larger than the population mean across different classes. The distribution greatness indicates that active compounds hitting targets in the same class are significantly more similar than the ones hitting targets from different classes. Figure 1B demonstrates the statistical significance of the distribution greatness of intra-class against inter-class.

Out of 645 target classes we collected in the classification of druggable human proteome, we elicited 66 classes of interest, which come from 11 upper-level classes. We focused on the target class pairs across which there are at least a finite number k of pairwise active compounds (k=4). By employing the same validation approach, we examined the relationship between molecular properties and target classes (see Figure 1D). For example, we evaluated the categorization of 1) three members of the GPCR superfamily, such as the rhodopsin-like GPCRs, the secretin-like GPCRs, and the metabotropic glutamate receptor family, 2) five ligand type-associated receptors, such as amine-associated receptors, amino acid receptors, lipid G protein–coupled receptors, nucleoside receptors, and peptide receptors, 3) the 13 GPCR sub families, 4) eight kinase groups, and 5) three ion channel families. A significant p-value reported in the unpaired two-sample one-tailed t-test tested our hypothesis. The assessment of five GPCR ligand types in DTO demonstrated that the categorization has been distinguished unambiguously (see Figure 1D); the assessment of 13 GPCR subfamily classification showed the same merit, except the neurotensin receptors and endothelin receptors; except the kinase TK group, other groups have not been differentiated consistently (see Figure 1D). The confusion may be caused by the innate structures of protein kinases since a large number of protein kinase enzymes share a common cofactor and similar three-dimensional structure of the catalytic site. Thus, active compounds that are screened may not effectively differentiate kinases. Another few of non-proved agreements may arise from the possible biased experimental selectivity of compounds, e.g., with the limited number of compound-GPCR combinations.

We further examined the relationship between target class and molecular properties of the related ligand sets, not individual compound but overall active compounds against targets. Our hypothesis is that overall related ligand sets active against intra-class are more similar than inter-class. Here we measured the target-to-target similarity in terms of the related ligand set similarity using the SEA method (see Methods). We implemented the SEA method in [5] and calculated the Z-score derived from the SEA underlying statistical model, which is

the normalized raw score summarized over the related ligand set(s) (see Methods). By employing unpaired two-sample one-tailed t-test, we measured the statistical significance of the greatness of the population mean of the Z-scores within target class against across different classes. For example, we have a total of 12 receptors in the 5-hydroxytryptamine receptor (5-HT receptors) GPCR subfamily since the diversity of metabotropic 5-HT receptors is increased by alternative splicing that produces isoforms of the 5-HT2A, 5-HT2C, 5-HT4, 5-HT6 and 5-HT7 receptors. We calculated the similarity of pairwise receptors in the 5-HT receptor subfamily by SEA. We also calculated the pairwise target similarity across the 5-HT receptor subfamily and one of any other 11 subfamilies of interest. After the t-test examination, we observed that targets in 5-HT receptor subfamily are more discriminatively different from other 10 GPCR subfamilies, but less distinctly different from dopamine receptors. Authors in [21] demonstrated that 5-HT receptors may modulate the release of striatal dopamine receptors in acute psychoses, which indicates that both subfamilies may function similar, especially in the acute psychotic states.

E. Promiscuity across different target families or a number of target subtypes

The existing and emerging bioassay activity/inhibition data offers a chance of exploring the pharmacology target space through the chemistry of their ligands. The SEA method used multivariate regression model to investigate pairwise target-to-target associations in terms of the related ligand set chemical structural similarity [5,6, 24-30]. We focused on 237 PDSP targets, and through the SEA method we calculated the similarity score of 56,169 target pairs on the basis of the similarity of related ligand sets, where no single ligand need be shared between pairwise targets, but overall. Through the method, we obtained 5870 pairs under the threshold E-value < 10-10 and built target-to-target relatedness network (see Figure 2 and Method). The smaller E-value, the higher confidence the pairwise targets could be related. In the network, each node is a drug target active against at least four compound and nodes are linked by an edge if their expected similarity E-value is satisfied with the specified threshold (see methods), indicating the relatedness of the associated targets. Connected components are formed and targets linked in a component are similar. The target-to-target relatedness network may be varied with the change of the threshold. The smaller the threshold, the more isolated components would be formed. However the target-to-target relatedness network built from the bioactivity data can indicate the diversity of targets inhibited by compounds through multiple distinct mechanisms, e.g., same or different subtypes or remotely related targets. For example, the NECA drug acts through blocking a number of ADORA1, ADORA2A, ADORA2B, and ADORA3 adenosine receptors.

Our observation is as follows. Some of them are consistently from the same IDG family and others are observed as the "cross-over" evidence that targets of interest are related to other targets coming from different types or subtypes of IDG families (see Figure 2). For instance, the glutamate receptor, ionotropic, AMPA 3 (GRIA3) is an ion channel and there are two GPCR targets (GRM2 and GRM3) and three other ion channels (GRIA1, GRIA2, and GRIK4) related to it. Among all those GPCR targets of endogenous ligand-binding mode, somatostatin receptors, prostaglandin receptors, galanin receptors, and adenosine receptors are discovered in the case of the mutual relationships, though coming from different GPCR families or sub-families. We found a list of 24 GPCR targets of different subtypes related to 5-HT3A receptor which is a kind of the serotonin receptors found in the central and peripheral nervous systems. The 5-HT3A receptor is closely related to other serotonin G protein-coupled receptors such as the 5-hydroxytryptamine subtypes 1, 2 and 4–7 (5-HT1,2,4–7) [5,7,8]. It is also related to targets that appear far from the serotonin receptors as shown in many bioinformatics metrics reflected from the ontology, but all of them have endogenous ligand-binding mode. These evidences may suggest that the promiscuity that small molecule compounds possessed is possibly across target boundaries. The polypharmacology across different target families or a number of target subtypes has been also pointed out in [5, 27, 32]. It turns out that local binding site may be more important than global structural or function information.

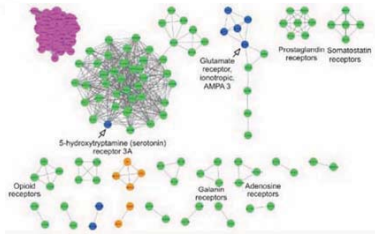


Figure 2. **Target-to-target relatedness network extracted from PDSP**, showing that these functional related ligand-focused targets are grouped in terms of ligand set similarity. Coloring represents different IDG (illuminating druggable genome) families. Red represents kinases, green GPCRs, blue ion channels, and orange nuclear receptors. We observed that completely connected graph have very similar gene symbols, for example, somatostatin receptors, prostaglandin receptors, opioid receptors, galanin receptors, and adenosine receptors. The completely connected components may be varied with the changing of the threshold (the lower the threshold, the more the isolated compounds would be formed). Most of connected components possess a unique color, which indicates that these targets in the component are from the same IDG family. We observed the "cross-over" evidence that interesting targets are related to targets of different types or subtypes of IDG families (see the arrowed nodes). For instance, the glutamate receptor, ionotropic, AMPA 3 (GRIA3) is an ion channel and there are two GPCR targets (GRM2 and GRM3) and three other ion channels (GRIA1, GRIA2, and GRIK4) related to it. 5-HT_{3A} receptor is an ion channel and a list of 24 GPCR targets is related to it.

III. DISCUSSION

We studied the pharmacology space problem in drug discovery by designing a gene family-led meta-level knowledge-driven approach to analyze drug-target data and examine the bias of target selection in experimental design. We analyzed the target promiscuity of compounds in PDSP drug screening data, designed and implemented a uniform way to visualize and compare the coverage of drug targets and to ensemble target-to-target similarity on the basis of associated ligand set similarity, in conjunction with knowledge represented in drug target ontology (DTO). Our ultimate goal

is to integrate and coordinate heterogeneous metadata for understanding dynamic biological systems and to develop improved computational tools that can quantify context-dependent disease-drugs-targets combinations for personalized medication. Such tools can potentially inform the prevention, diagnosis and treatment of diseases.

During this analysis, we note that different types of bias may exist in a binding bioactivity data. One possible concern is the chemical space bias arising from the studied ligands or screening libraries toward biogenic molecules [30]. Another concern exists in the selection of a panel of protein targets against a given compound. We note that some proteins are never profiled against a drug in both screening data, for example, all 424 olfactory receptor proteins. It seems experimentally reasonable since there is the lack of comprehensive understanding of olfactory mechanism in relation to diverse chronic psychiatric diseases (the sense of smells is neglected by clinicians in routine work) [57]. However, it is noticed that olfaction plays an important role in protecting us from environmental dangers. Olfactory dysfunction is common in later life and early warning of neurodegenerative disease, which is receiving more attention as a sign in the early detection and differential diagnosis of neurological and mental diseases. The timely effective treatment of olfactory loss is facilitative to cure related complications of neurodegenerative disease.

It is observed that the selectivity of most drugs is usually profiled against a panel of same kinase families or GPCRs families. It is reasonable since "binding sites within members of conserved and evolutionarily related targets are generally conserved and prone to multi-target inhibition" [4]. However false negatives may be expected since not all targets have been profiled for a given compound. Especially, it is rarely profiled for those targets belonging to different gene families, distantly related or unrelated. Those unintended off-targets may cause unpredicted side effects or unforeseen serious toxicity. It is actually impossible to specify all multi-targets by prior knowledge if the mechanism of drug action is not clear. It is also currently infeasible to profile all drugs against a panel of all proteins for diseases at different phrases, disease progression, and traits that affects specific tissues, tissue compartments, or cells. Computational tools therefore offer an attractive alternative, especially if they can be shown to reduce the search space or leverage the priority of potential drug-specific target combinations in the context since the search space for multi-targets and multi-drugs is huge.

ACKNOWLEDGMENT

We especially thank the investigators and their colleagues who generated and generously deposited drug-screening data. This research was supported by grant U54HL127624 awarded by the National Heart, Lung, and Blood Institute through funds provided by the trans-NIH Library of Integrated Network-based Cellular Signatures (LINCS) Program (http://www.lincsproject.org/) and the trans-NIH Big Data to Knowledge (BD2K) initiative (http://www.bd2k.nih.gov).

REFERENCES

[1] K., W. K., S., M. F., Huang, X. P., Lansu, K., McCorvy, J. D., Giguère, P. M., ... & Roth, B. L. (2015). PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. Nature structural & molecular biology, 22(5), 362-369.

[2] Besnard J, Ruda GF, Setola V, Abecassis K, Rodriguiz RM, Huang XP, Norval S, Sassano MF, Shin AI, Webster LA, Simeons FR, Stojanovski L, Prat A, Seidah NG, Constam DB, Bickerton GR, Read KD, Wetsel WC, Gilbert IH, Roth BL, Hopkins AL. (2012) Automated design of ligands to polypharmacological profiles. Nature. 2012 Dec 13;492(7428):215-20. doi: 10.1038/nature11691.

[3] Anton Yuryev (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Pathway Analysis for Drug Discovery. ISBN 978-0-470-10705-8

[4] Guilio Rastelli and Luca Pinzi. (2015) Computational polypharmacology comes of age. Frontiers in Pharmacology. Volume 6 | article 157. doi: 10.3389/fphar.2015.00157

[5] Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. (2007) Relating protein pharmacology by ligand chemistry. Nat Biotech 25 (2), 197-206.

[6] Hert, J., Keiser, M. J., Irwin, J. J., Oprea, T. I., & Shoichet, B. K. (2008). Quantifying the relationships among drug classes. Journal of chemical information and modeling, 48(4), 755-765.

[7] Roth, B.L., Sheffler, D.J. & Kroeze, W.K. (2004) Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. Nat. Rev. Drug Discov. 3, 353–359.

[8] Kroeze, W.K., Kristiansen, K. & Roth, B.L. (2002) Molecular biology of serotonin receptors structure and function at the molecular level. Curr. Top. Med. Chem. 2, 507–528.

[9] Wesley K. Kroeze and Bryan L. Roth. (2012) Polypharmacological Drugs: " Magic Shotguns " for Psychiatric Diseases. Peters, Jens-Uwe, ed. Polypharmacology in Drug Discovery. Somerset, NJ, USA: John Wiley & Sons.

[10] Uhlén M et al, 2015. Tissue-based map of the human proteome. Science

[11] Piñero, J., Queralt-Rosinach, N., Bravo, À., Deu-Pons, J., Bauer-Mehren, A., Baron, M., ... & Furlong, L. I. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database, 2015, bav028.

[12] Pfam database.

[13] Mestres, Jordi, et al. (2008) Data completeness—the Achilles heel of drug-target networks. Nature biotechnology 26.9: 983-984.

[14] GO and related tools.

[15] Zhidian Du, Lin Li, Chin-Fu Chen, Philip. S. Yu, and James Z. Wang. (2009) G-sesame: web tools for go-term-based gene similarity analysis and knowledge discovery. Nucleic Acids Research, 37:W345–W349

[16] Li H, Pellegrini M, Eisenberg D. (2005) Detection of parallel functional modules by comparative analysis of genome sequences. Nat Biotechnol 2005, 23(2):253-260.

[17] Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006 Jan 1;34(Database issue):D668-72. 16381955

[18] Uhlén et al (2015). Tissue-based map of the human proteome. Science PubMed: 25613900 DOI: 10.1126/science.1260419

[19] Stephan C. Schürer and Steven M. Muskal (2012) Kinome-wide Activity Modeling from Diverse Public High-Quality Data Sets. J Chem Inf Model. 2013; 53(1): 27–38 DOI: 10.1021/ci300403k

[20] Law V et al. (2014) DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. 2014 Jan 1;42(1):D1091-7. PubMed ID: 24203711

[21] Vollenweider, F. X., Vontobel, P., Hell, D., & Leenders, K. L. (1999) 5-HT modulation of dopamine release in basal ganglia in psilocybin-induced psychosis in man—a PET study with [11C] raclopride. Neuropsychopharmacology, 20(5), 424-433.

[22] A.P. Bento, A. Gaulton, A. Hersey, L.J. Bellis, J. Chambers, M. Davies, F.A. Kruger, Y. Light, L.Mak, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos and J.P. Overington (2014) The ChEMBL

[23] LINCS.

[24] Lounkine, E., Keiser, M. J., Whitebread, S., Mikhailov, D., Hamon, J., Jenkins, J., ... Urban, L. (2012). Large Scale Prediction and Testing of Drug Activity on Side-Effect Targets. Nature, 486(7403), 361–367.

[25] Laggner, C., Kokel, D., Setola, V., Tolia, A., Lin, H., Irwin, J. J., ... Shoichet, B. K. (2011). Chemical Informatics and Target Identification in a Zebrafish Phenotypic Screen. Nature Chemical Biology, 8(2), 144–146.

[26] Yadav, P. N., Abbas, A. I., Farrell, M. S., Setola, V., Sciaky, N., Huang, X.-P., ... Roth, B. L. (2011). The Presynaptic Component of the Serotonergic System is Required for Clozapine's Efficacy. Neuropsychopharmacology, 36(3), 638–651.

[27] Keiser, M. J., Irwin, J. J., & Shoichet, B. K. (2010). The Chemical Basis of Pharmacology. Biochemistry, 49(48), 10267–10276.

[28] DeGraw, A. J., Keiser, M. J., Ochocki, J. D., Shoichet, B. K., & Distefano, M. D. (2010). Prediction and evaluation of protein farnesyltransferase inhibition by commercial drugs. Journal of Medicinal Chemistry, 53(6), 2464–2471.

[29] Keiser, M. J., Setola, V., Irwin, J. J., Laggner, C., Abbas, A., Hufeisen, S. J., ... Roth, B. L. (2009). Predicting new molecular targets for known drugs. Nature, 462(7270), 175–181.

[30] Hert, J., Irwin, J. J., Laggner, C., Keiser, M. J., & Shoichet, B. K. (2009). Quantifying Biogenic Bias in Screening Libraries. Nature Chemical Biology, 5(7), 479–483.

[31] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., ... Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research, 13(11), 2498–2504.

[32] Paolini, G. V., Shapland, R. H., van Hoorn, W. P., Mason, J. S., & Hopkins, A. L. (2006). Global mapping of pharmacological space. Nature biotechnology, 24(7), 805–815.

[33] Yıldırım, M. A., Goh, K. I., Cusick, M. E., Barabási, A. L., & Vidal, M. (2007). Drug—target network. Nature biotechnology, 25(10), 1119-1126.

[34] Chen B, Ding Y, Wild DJ (2012) Assessing Drug Target Association Using Semantic Linked Data. PLoS Comput Biol 8(7): e1002574. doi:10.1371/journal.pcbi.1002574

[35] A. J. Butte, I. S. Kohane (2000) Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. Pac Symp Biocomput. 2000:418-29 PMID: 10902190

[36] Bauer-Mehren,A., Rautschka,M., Sanz,F. et al. (2010) DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. Bioinformatics, 26, 2924–2926

[37] Kim et al.(2014) A draft map of the human proteome. Nature. 509, 575-581.

[38] Schriml, L. M., Arze, C., Nadendla, S., Chang, Y.-W. W., Mazaitis, M., Felix, V., ... Kibbe, W. A. (2012) Disease Ontology: a backbone for disease semantic integration. Nucleic Acids Research, 40(Database issue), D940–D946.

[39] Bauer-Mehren,A., Bundschus,M., Rautschka,M. et al. (2011) Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases. PLoS One, 6, e20284

[40] Tang J, Karhinen L, Xu T, Szwajda A, Yadav B, Wennerberg K, et al. (2013) Target Inhibition Networks: Predicting Selective Combinations of Druggable Targets to Block Cancer Survival Pathways. PLoS Comput Biol 9(9): e1003226. doi:10.1371/journal.pcbi.1003226

[41] Huang,L., Li,F., Sheng,J. et al. (2014) DrugComboRanker: drug combination discovery based on target network analysis. Bioinformatics, 30, i228–i236.

[42] Liu, Y., Wei, Q., Yu, G., Gai, W., Li, Y., & Chen, X. (2014) DCDB 2.0: a major update of the drug combination database. Database, 2014, bau124.

[43] Knight, Z. A., Lin, H., & Shokat, K. M. (2010) Targeting the cancer kinome through polypharmacology. Nature Reviews Cancer, 10(2), 130-137.

[44] Cheng, Q., Ursu, O., Oprea, T. I., & Schurer, S. (2017). Learning reference-enriched approach towards large scale active ontology alignment and integration. BIBM 2017 pp. 1658-1663.

[45] Jalencas, X., & Mestres, J. (2013) Identification of Similar Binding Sites to Detect Distant Polypharmacology. Molecular Informatics, 32(11-12), 976-990.

[46] Miller, D. W., & Dill, K. A. (1997) Ligand binding to proteins: the binding landscape model. Protein science, 6(10), 2166-2179.

[47] Cheng, F., Liu, C., Jiang, J., Lu, W., Li, W., Liu, G., ... & Tang, Y. (2012) Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS Comput Biol, 8(5), e1002503.

[48] Campillos M et al. (2008) Drug target identification using side-effect similarity. Science 321:263–266

[49] Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M (2008) Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics 24: i232–240.

[50] Yamanishi, Y., Kotera, M., Kanehisa, M., & Goto, S. (2010) Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. Bioinformatics, 26(12), i246-i254.

[51] Santos, R., Ursu, O., Gaulton, A., Bento, A.P., Donadi, R.S., Bologa, C., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T.I., & Overington, J.P. A Comprehensive Map of Molecular Drug Targets. Nature Rev. Drug Discov. doi:10.1038/nrd.2016.230 (2016)

[52] Schadt, E. E., Friend, S. H., & Shaywitz, D. A. (2009) A network view of disease and compound screening. Nature reviews Drug discovery, 8(4), 286-295.

[53] Liu, W., Wu, A., Pellegrini, M., & Wang, X. (2015) Integrative analysis of human protein, function and disease networks. Scientific reports, 5.

[54] Hopkins, A. L. (2008) Network pharmacology: the next paradigm in drug discovery. Nature chemical biology, 4(11), 682-690.

[55] Overington, J. P., Al-Lazikani, B., & Hopkins, A. L. (2006) How many drug targets are there?. Nature reviews Drug discovery, 5(12), 993-996.

[56] Larance, M., & Lamond, A. I. (2015) Multidimensional proteomics for cell biology. Nature Reviews Molecular Cell Biology.

[57] Kivity S., Ortega-Hernandez O. D. and Shoenfeld Y. (2009) Olfaction--a window to the mind. Isr Med Assoc J. 2009 Apr;11(4):238-43. PMID: 19603599

[58] Przydzial, M. J., Bhhatarai, B., Koleti, A., Vempati, U., & Schürer, S. C. (2013). GPCR ontology: development and application of a G protein-coupled receptor pharmacology knowledge framework. Bioinformatics. 2013 Dec 15;29(24):3211-9. doi: 10.1093/bioinformatics/btt565. Epub 2013 Sep 29.

[59] Lin, Y., Mehta, S., McGinty, H. K., Turner, J. P., Vidovic, D., Forlin, M., ... & Mathias, S. L. (2017). Drug Target Ontology to Classify and Integrate Drug Discovery Data. bioRxiv, 117564

[60] Cannon, D.C., Yang, J.J., Mathias, S.L., Ursu, O., Mani, S., Waller, A., Schürer, S.C., Jensen, L.J., Sklar, L.A., Bologa, C.G. and Oprea, T.I., 2017. TIN-X: target importance and novelty explorer. Bioinformatics, p.btx200.

[61] Ursu, O., Holmes, J., Knockel, J., Bologa, C.G., Yang, J.J., Mathias, S.L., Nelson, S.J. and Oprea, T.I., 2017. DrugCentral: online drug compendium. Nucleic acids research, 45(D1), pp.D932-D939

[62] Loging WT, editor. Bioinformatics and Computational Biology in Drug Discovery and Development. Cambridge University Press; 2016 Mar 17.

[63] Acharya, A. and Garg, L.C., 2016. Drug Target Identification and Prioritization for Treatment of Ovine Foot Rot: An In Silico Approach. International journal of genomics, 2016

[64] Russ, A. P., & Lampel, S. (2005). The druggable genome: an update. Drug discovery today, 10(23), 1607-1610.

[65] Kolaja, K. (2012). Drug discovery: computer model predicts side effects. Nature, 486(7403), 326-328.

[66] Lounkine, E., Keiser, M.J., Whitebread, S., Mikhailov, D., Hamon, J., Jenkins, J., Lavan, P., Weber, E., Doak, A.K., Côté, S. and Shoichet, B.K., 2012. Large scale prediction and testing of drug activity on side-effect targets. Nature, 486(7403), p.361.

[67] Kobourov, S. G. (2004). Force-directed drawing algorithms.

[68] Cheng, Q., Lopez, F. A., Duran, C., Camarillo, C., Oprea, T. I., & Schurer, S. (2017). The ontology reference model for visual selectivity analysis in drug-target interactions. In 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 2091-2097). IEEE.

Theory in Biosciences
https://doi.org/10.1007/s12064-019-00275-5

ORIGINAL ARTICLE

A mathematical and numerical study of a SIR epidemic model with time delay, nonlinear incidence and treatment rates

Kanica Goel[1] · **Nilam**[1]

Received: 2 May 2018 / Accepted: 8 January 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract
A novel nonlinear time-delayed susceptible–infected–recovered epidemic model with Beddington–DeAngelis-type incidence rate and saturated functional-type treatment rate is proposed and analyzed mathematically and numerically to control the spread of epidemic in the society. Analytical study of the model shows that it has two equilibrium points: disease-free equilibrium (DFE) and endemic equilibrium (EE). The stability of the model at DFE is discussed with the help of basic reproduction number, denoted by R_0, and it is shown that if the basic reproduction number R_0 is less than one, the DFE is locally asymptotically stable and unstable if R_0 is greater than one. The stability of the model at DFE for $R_0 = 1$ is analyzed using center manifold theory and Castillo-Chavez and Song theorem which reveals a forward bifurcation. We also derived the conditions for the stability and occurrence of Hopf bifurcation of the model at endemic equilibrium. Further, to illustrate the analytical results, the model is simulated numerically.

Keywords Epidemic model · Beddington–DeAngelis-type incidence rate · Saturated treatment rate · Stability · Bifurcation · Center manifold theory

Mathematics Subject Classification 34D20 · 92B05 · 37M05

Introduction

Mathematical analysis and modeling of infectious diseases play a crucial role in studying a wide range of infectious diseases to better understand the transmission dynamics and have the capacity to influence expectations and to decide and assess control strategies. Elementary descriptions of infectious diseases have been considered mainly by three epidemiological classes which measure the susceptible portion of population, the infected and the removed (recovered) ones. Authors around the world have proposed different kinds of epidemic models such as susceptible–infected (SI) (Mukherjee 1996), susceptible–infected–susceptible (SIS) (Hethcote and van den Driessche 1995), susceptible–infected–recovered (SIR) (Kumar and Nilam 2018a), susceptible–infected–recovered–susceptible (SIRS) (Mena-Lorca and Hethcote 1992), susceptible–exposure–infected–recovered (SEIR) (Dubey et al. 2013; Tipsri and Chinviriyasit 2014), susceptible–Vaccinated–exposure–infected–recovered (SVEIR) (Gumel et al. 2007) and many more, to understand the dynamics of disease transmission. In the mathematical epidemiological literature, several authors have studied the epidemiological models with latent or incubation period, because many diseases have a latent or incubation period, during which the susceptible individual becomes infected but is not yet infectious. Such latency in disease transmission can be modeled by a delay differential equation. Delay differential equations (DDE) have been a very successful tool to capture the effect of varying infectious period in a range of SIR, SIS, SIRS and other epidemic models. Hethcote and van den Driessche (1995) have studied an SIS epidemic model with constant time delay, which accounts for duration of infectiousness. Also, Song and Cheng (2005) have studied the impact of time delay on the stability of the positive equilibrium, as a resultant of which conditions have been stated for the asymptotical stability of the endemic equilibrium for all delays. Xu and Ma (2009b), Hattaf et al. (2013), Kumar and Nilam (2018a, b) considered the effect of time delay on SIRS and SIR models, respectively, and provided the conditions for the stability of their proposed models.

The incidence rate of a disease is the number of new cases per unit time and plays a crucial role in the study of transmission of disease dynamics. Several authors suggested different types of incidence rate. Firstly, the bilinear incidence rate βSI (Kermack and McKendrick 1927; Bailey 1975; Anderson and May 1992; Brauer and Castillo-Chavez 2001; Zhang and Suo 2010) is based on the law of mass action, which describes the situation that if the number of susceptibles increases, the number of individuals who become infected per unit of time increases, which is not realistic. In reality, however, by the impact of media, open mindfulness or individual experience, people apply careful steps that decrease the contact number or the transmission potential. Since nonlinearity in the incidence rates has been seen in disease transmission dynamics, it has been proposed that the standard bilinear incidence rate shall be modified into a nonlinear incidence rate by numerous authors (Capasso and Serio 1978; Korobeinikov 2007). Several authors such as Anderson and May (1978), Wei and Chen (2008), Zhang et al. (2008), Li et al. (2009), Li and Muldowney (1995), Korobeinikov and Maini (2005), Xu and Ma (2009b), Capasso and Serio (1978) suggested different types of nonlinear incidence rates, and they incorporated nonlinear-type incidence rate in their model and studied the disease dynamics. Beddington (1975) and DeAngelis et al. (1975) independently introduced nonlinear incidence rate known as Beddington–DeAngelis-type incidence rate. Later, some authors (Kaddar 2010; Elaiw and Azoz 2013; Dubey et al. 2015) used this incidence rate in epidemiological models. To contribute to the nonlinear dynamics of infectious disease, in the present study we introduce the incidence rate as Beddington–DeAngelis type. Since nonlinear-type incidence rate alone cannot determine the actual transmission of the disease dynamics completely, for a more realistic model, inclusion of time delay must be considered. Therefore, in the present study we are incorporating the effect of time lag into Beddington DeAngelis functional-type incidence rate and investigating its impact on the disease dynamics.

It is well known that to prevent and control the spread of epidemics, the treatment rate has played a substantial role. In the classical epidemic model, treatment rate was considered to be either constant (Wang and Ruan 2004) or proportional to number of infected individuals (Wang 2006). This type of treatment rate is suitable in the case when the number of infectives is small and treatment resources are sufficient and unsuitable when the number of the infectives is large and treatment resources are limited. To control the disease, most researchers focus on a nonlinear-type treatment rate. Dubey et al. (2013, 2015, 2016) introduced the nonlinear treatment rate as Holling type II, Holling type III and Holling type IV in their model and proposed nonlinear dynamics to control the epidemic. Holling type III defines the condition in which removal rate initially becomes quick with increment in infectives and after that it develops gradually and lastly settles down to maximum saturated value. After this any increase in infectives won't influence the removal rate (Dubey et al. 2013). Motivated by these work, in the present study we take the nonlinear treatment rate as saturated treatment rate functional type of the form

$$g(I) = \frac{aI^2}{bI^2 + cI + 1} \quad \text{(where } I \geq 0,\ a, b > 0 \text{ and } c \geq 0\text{)}$$

In the present study our aim is to understand and predict actual transmission of the infectious diseases and investigate the effect of treatment rate to provide effective control strategy that can hasten the dying out of disease. Therefore, motivated by the work of (Dubey et al. 2013, 2015, 2016) we consider a susceptible–infected–recovered (SIR) epidemic model setting with a nonlinear Beddington–DeAngelis functional-type incidence rate that includes awareness/inhibitions behavior by susceptibles and infectives with the inclusion of time lag (describing the latent period). We also consider nonlinear saturated treatment rate to provide a control strategy of the infectious diseases and show that if this behavioral response is delayed, then this simple looking system can exhibit interesting dynamics. For the dynamics of the model, we discuss the existence and stability behavior of model for equilibriums. The existence of a Hopf bifurcation is also discussed.

The rest of the paper is organized as follows: SIR model for disease dynamics and the basic mathematical properties of the model are studied in "Mathematical framework" section. In "Equilibrium and stability analysis" section, the basic reproduction number is derived, and we study the qualitative behavior of the model via stability at the equilibrium points and Hopf bifurcation when time delay is considered as the bifurcation parameter. Furthermore, in "Numerical simulation" section, numerical simulations are presented using MATHEMATICA 11, and lastly, we present our conclusion in "Conclusion" section.

Mathematical framework

In this section we present our mathematical model which is described by delay differential equation constructed from the interaction among the compartments: susceptible, infective and recovered for the model equation.

✉ Nilam
rathi.nilam@gmail.com

Kanica Goel
kanica.dtu2016@gmail.com

[1] Department of Applied Mathematics, Delhi Technological University, Delhi 110042, India

Published online: 21 January 2019

The basic model equations

We consider that in the region under consideration, the total population is $N(t)$ at time t. The total population $N(t)$ is divided into three subclasses: the susceptible class $S(t)$, the infective $I(t)$ and the recovered individuals $R(t)$. Susceptible class $S(t)$ consists of those individuals who are healthy but can contract the disease. The infective class $I(t)$ is the class of individuals who have contracted the disease and are now infectious and can transmit the disease to others. $R(t)$ represents the number of individuals who are recovered from the diseases with partial immunity or treatment. The conceptual description of the model is represented by a block diagram (Fig. 1).

The model we present in this paper is defined by the following system of nonlinear delay differential equations:

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= \pi - \mu S - \frac{\beta S(t-\tau)I(t-\tau)}{1+\alpha S(t-\tau)+\gamma I(t-\tau)},\\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{\beta S(t-\tau)I(t-\tau)}{1+\alpha S(t-\tau)+\gamma I(t-\tau)} - (\mu+d+\theta)I - \frac{aI^2}{bI^2+cI+1},\\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \theta I + \frac{aI^2}{bI^2+cI+1} - \mu R.
\end{aligned} \tag{1}$$

where $\tau > 0$ represents the time delay describing the latent period of the disease.

The initial conditions $\phi = (\phi_1, \phi_2, \phi_3)$ of (1) are defined in the Banach space

$$\begin{aligned}
C_+ = \big\{\phi \in C\big([-\tau, 0], R_+^3\big) : \phi_1(\Omega) = S(\Omega),\\
\phi_2(\Omega) = I(\Omega), \quad \phi_3(\Omega) = R(\Omega)\big\},
\end{aligned}$$

where $R_+^3 = \{(S, I, R) \in R^3 : S \geq 0, I \geq 0, R \geq 0\}$. Biologically, we assume that $\phi_i > 0 \quad (i = 1, 2, 3)$.

In the above model (1), we consider the total population $N(t)$ at time t, with immigration of susceptible population with a constant rate π. The parameters μ, d and θ represent the natural death rate, disease-induced death rate and recovery rate, respectively. In epidemiological models, by modifying the incidence rate (the number of individuals who become infected per unit of time), behavioral changes during an epidemic can be seen to shape the dynamics very differently from the kind predicted by simple models. Therefore, in the present model (1), we have taken the incidence rate as Beddington–DeAngelis type with delay:

$$f(S, I) = \frac{\beta S(t-\tau)I(t-\tau)}{1+\alpha S(t-\tau)+\gamma I(t-\tau)}$$

Here, β is the effective contact rate (or transmission rate) of susceptibles with infectives, α is a measure of inhibition effect, such as preventive measures taken by susceptible individuals, and γ is a measure of inhibition effect such as treatment with respect to infectives. The incidence function f includes some special cases. For instance, if we set $\alpha = 0, \gamma = 0$, then the incidence rate is bilinear (Gumel et al. 2007) and if $\alpha = 0$ the incidence rate, describing saturated effects of the prevalence of infectious diseases, is used in McCluskey (2010) and Xu and Ma (2009a). The term $g(I) = \frac{aI^2}{bI^2+cI+1}$ in model (1) is the saturated treatment rate, where $a > 0$ represents the treatment rate of infected individuals (cure rate), $b > 0$ is the limitation rate in treatment availability, and $c \geq 0$ is the saturation constant in absence of inhibitory effect.

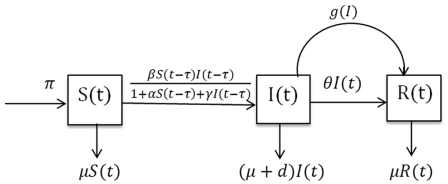


Fig. 1 Transfer diagram for the delayed SIR model with the susceptible class $S(t)$, the infective class $I(t)$ and the recover class $R(t)$

Basic properties of the model

The equations of the model (1) monitor populations. From the proposition 2.1 in Hattaf et al. (2013) and proposition 2.3. in Yang and Sun (2015), it can be shown that all state variables of the model (1) are nonnegative. That is, $(S, I, R) \in R_+^3$. Also, for ecological reasons, we supposed that all parameters π, μ, β, α, γ, d, θ, a, b are positive and c is nonnegative.

Since $N(t) = S(t) + I(t) + R(t)$, the governing equations of model (1) can be rewritten as,

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \pi - \mu N - dI \tag{2}$$

$$\leq \pi - \mu N \tag{3}$$

Lemma 1 *All solutions of the model* (1) *starting in* R_+^3 *are bounded and eventually enter a compact attracting set*

$$\Phi = \left\{(S, I, R) \in R_+^3 : S(t) + I(t) + R(t) = N(t) \leq \frac{\pi}{\mu}\right\}.$$

Proof Continuity of the right-hand side of the model (1) and its derivative assure the well-posedness of the model for $N(t) > 0$. The invariant region for the existence of the solutions can be determined as given below:

$$0 < \liminf_{t\to\infty} N(t) \leq \limsup_{t\to\infty} N(t) \leq \frac{\pi}{\mu}$$

Since $N(t) > 0$ on $[-\tau, 0]$ by assumption, $N(t) > 0$ for all $t \geq 0$. Therefore, with the help of Eq. (3), it can be seen that for any finite time t, $N(t)$ cannot blow up to infinity. The model system is dissipative (solutions are bounded), and consequently, the solution exists globally for all $t > 0$ in the invariant and compact set

$$\Phi = \left\{(S, I, R) \in R_+^3 : S(t) + I(t) + R(t) = N(t) \leq \frac{\pi}{\mu}\right\},$$

As $N(t)$ approaches zero, $S(t)$, $I(t)$, and $R(t)$ also approach zero. Thus, each of these terms tends toward zero as $N(t)$ does. Thus, it is reasonable to interpret these terms as zero when $N(t) = 0$. □

Remark 1 In this region Φ, basic results such as usual local existence, uniqueness and continuation of solutions are valid for model (1). Hence, there exists a unique solution $(S(t), I(t), R(t))$ of model (1) starting in the interior of Φ that exists on a maximal interval $[0, \infty)$ if solutions remain bounded (Kuang 1993).

Equilibrium and stability analysis

From the model (1) we infer that since $R(t)$ does not appear in equations for $\frac{\mathrm{d}S}{\mathrm{d}t}$ and $\frac{\mathrm{d}I}{\mathrm{d}t}$, it is sufficient to analyze the behaviors of solutions of (1) by the following system of DDEs:

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= \pi - \mu S - \frac{\beta S(t-\tau)I(t-\tau)}{1+\alpha S(t-\tau)+\gamma I(t-\tau)}\\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{\beta S(t-\tau)I(t-\tau)}{1+\alpha S(t-\tau)+\gamma I(t-\tau)} - (\mu+d+\theta)I\\
&\quad - \frac{aI^2}{bI^2+cI+1}
\end{aligned} \tag{4}$$

with initial conditions $\phi = (\phi_1, \phi_2)$ of (4) defined in the Banach space

$$C_+ = \big\{\phi \in C\big([-\tau, 0], R_+^2\big) : \phi_1(\Omega) = S(\Omega), \phi_2(\Omega) = I(\Omega)\big\},$$

where $R_+^2 = \{(S, I) \in R^2 : S \geq 0, I \geq 0\}, \phi_i > 0 \quad (i = 1, 2)$.

Now, we analyze the equilibrium points of the system (4). There are only two types of physically, and in addition biologically, relevant equilibria, namely

1. $E_0(S_0, 0) = E_0\left(\frac{\pi}{\mu}, 0\right)$, disease-free equilibrium (DFE).
2. $E_e(S^*, I^*)$, positive or endemic equilibrium (EE)

where S^* and I^* are given in Section 3.2.

Disease-free equilibrium and its stability

System (4) has always a disease-free equilibrium point of the form $E_0\left(\frac{\pi}{\mu}, 0\right)$ (that is, there is no infection present in the community and all individuals are susceptible) which is obtained by setting right-hand sides of the system (4) to zero. The characteristic equation of the linearization of model (4) near the disease-free equilibrium $E_0\left(\frac{\pi}{\mu}, 0\right)$ is given by

$$(\mu+\lambda)\left(\frac{\beta\pi e^{-\lambda\tau}}{\mu+\alpha\pi} - (\mu+d+\theta) - \lambda\right) = 0 \tag{5}$$

One of the roots of Eq. (5) is given by $\lambda_1 = -\mu$ and other root can be obtained from

$$\frac{\beta\pi e^{-\lambda\tau}}{\mu+\alpha\pi} - (\mu+d+\theta) - \lambda = 0$$

The term $\frac{\beta\pi}{(\mu+\alpha\pi)(\mu+d+\theta)}e^{-\lambda\tau}$ at $\tau = 0$ is termed as basic reproduction number. In epidemiological research, the basic reproduction number, denoted by R_0, is defined as the average number of secondary infections caused by a single infected agent, during his/her entire infectious period, in a completely susceptible population (Van den Driessche and Watmough 2002). Thus, basic reproduction number for our model is given by

$$R_0 = \frac{\beta\pi}{(\mu+\alpha\pi)(\mu+d+\theta)}.$$

Analysis for $R_0 \neq 1$

Clearly, Eq. (5) has one negative root $\lambda_1 = -\mu$ and other root is obtained by the equation

$$\lambda + \mu + d + \theta - \frac{\beta\pi}{(\mu+\alpha\pi)}e^{-\lambda\tau} = 0$$

Let

$$\begin{aligned}
f(\lambda) &= \lambda + \mu + d + \theta - \frac{\beta\pi}{(\mu+\alpha\pi)}e^{-\lambda\tau}\\
&= \lambda + (\mu+d+\theta)\big(1 - R_0 e^{-\lambda\tau}\big)
\end{aligned}$$

If $R_0 > 1$, it is readily seen that for real λ,

$$f(0) = \lambda + (\mu+d+\theta)\big(1 - R_0\big) < 0, \quad \lim_{\lambda\to\infty} f(\lambda) = +\infty$$

Hence, $f(\lambda) = 0$ and $f'(\lambda) > 0$, so $f(\lambda) = 0$ has a unique positive real root if $R_0 > 1$.

If $R_0 < 1$, we assume that $Re\lambda \geq 0$.

We notice that

$$\begin{aligned}
Re\lambda &= \frac{\beta\pi e^{-(Re\lambda)\tau}\cos(Im\lambda)\tau}{(\mu+\alpha\pi)} - (\mu+d+\theta)\\
&< \frac{\beta\pi}{(\mu+\alpha\pi)} - (\mu+d+\theta) < 0
\end{aligned}$$

which contradicts our assumption. Therefore, if $R_0 < 1$ then λ will be a root of Eq. (5) with negative real part. The result can be written in form of the theorem which is given below:

Theorem 1 *The disease-free equilibrium E_0 is locally asymptotically stable if $R_0 < 1$ and unstable if $R_0 > 1$.*

Analysis at $R_0 = 1$

In this section, we analyze the behavior of system (4) when the basic reproduction number R_0 is equal to one. We observe that the Jacobian matrix of system (4) evaluated at $R_0 = 1$ and bifurcation parameter $\beta = \beta^* = \frac{(\mu+\alpha\pi)(\mu+d+\theta)}{\pi}$ has a simple zero eigenvalue and another eigenvalue with negative real part. Since linearization is not suitable to analyze the stability behavior of equilibrium points at $R_0 = 1$, center manifold theory (Sastry 1999) is used. Following assumptions are made to apply center manifold theory to system (4): let $S = x_1$ and $I = x_2$, then the system (4) can be written as

$$\begin{aligned}\frac{dx_1}{dt} &= \pi - \mu x_1 - \frac{\beta x_1(t-\tau)x_2(t-\tau)}{1+\alpha x_1(t-\tau)+\gamma x_2(t-\tau)} \equiv f_1\\ \frac{dx_2}{dt} &= \frac{\beta x_1(t-\tau)x_2(t-\tau)}{1+\alpha x_1(t-\tau)+\gamma x_2(t-\tau)} - (\mu+d+\theta)x_2\\ &\quad - \frac{ax_2^2}{{x_2}^2+bx_2+c} \equiv f_2\end{aligned} \tag{6}$$

Let J denotes the Jacobian matrix evaluated at $R_0 = 1$, and $\beta = \beta^*$ then

$$J = \begin{bmatrix} -\mu & -\frac{\beta^*\pi}{(\mu+\alpha\pi)} \\ 0 & 0 \end{bmatrix}$$

Let $w = \begin{bmatrix} w_1 & w_2 \end{bmatrix}$ and $u = \begin{bmatrix} u_1 & u_2 \end{bmatrix}^T$ be the left and right null-vectors of J corresponding to the zero eigenvalue. Then we have

$$w_1 = 0, \quad w_2 = 1 \quad \text{and} \quad u_1 = -\frac{\beta^*\pi}{\mu(\mu+\alpha\pi)}, \quad u_2 = 1$$

The nonzero partial derivatives associated with the functions f_1 and f_2 of the system (6) evaluated at $R_0 = 1$ and $\beta = \beta^*$ are

$$\left(\frac{\partial^2 f_2}{\partial x_2^2}\right)_{E_0} = -2a - \frac{2\pi\mu\beta\gamma}{(\mu+\alpha\pi)^2}, \quad \left(\frac{\partial^2 f_2}{\partial x_1 \partial x_2}\right)_{E_0} = \left(\frac{\partial^2 f_2}{\partial x_2 \partial x_1}\right)_{E_0} = -\frac{\pi\alpha\beta\mu}{(\mu+\alpha\pi)^2} + \frac{\beta\mu}{\mu+\alpha\pi} \quad \text{and}$$

$$\left(\frac{\partial^2 f_2}{\partial x_2 \partial \beta^*}\right)_{E_0} = \frac{\pi}{\mu+\alpha\pi}.$$

Then from Theorem 4.1 of Castillo-Chavez and Song (2004), the bifurcation constants a_1 and b_1 are calculated as follows:

$$\begin{aligned} a_1 &= \sum_{k,i,j=1}^{2} w_k u_i u_j \left(\frac{\partial^2 f_k}{\partial x_i \partial x_j}\right)_{E_0} \\ &= -\frac{2\beta^*\pi}{(\mu+\alpha\pi)}\left[-\frac{\pi\alpha\beta^*}{(\mu+\alpha\pi)^2} + \frac{\beta^*}{\mu+\alpha\pi}\right] - 2a - \frac{2\mu\pi\beta^*\gamma}{(\mu+\alpha\mu)^2} \\ &= -\frac{2\beta^*\pi\mu}{\mu+\alpha\pi}\left(\frac{\beta^*+\pi(\mu+\alpha\pi)}{(\mu+\alpha\pi)^2}\right) - 2a \\ &< 0 \quad \text{(as all parameters are positive).} \end{aligned}$$

and

$$b_1 = \sum_{k,i=1}^{2} w_k u_i \left(\frac{\partial^2 f_k}{\partial x_i \partial \beta^*}\right)_{E_0} = w_2 u_2 \left(\frac{\partial^2 f_2}{\partial x_2 \partial \beta^*}\right)_{E_0} = \frac{\pi}{\mu+\alpha\pi} > 0.$$

Sign of a_1 determines the nature of the bifurcation at $R_0 = 1$. The local analysis of the center manifold yields a parameter, a_1, whose sign indicates the existence and stability of a branch of endemic equilibria near the threshold $R_0 = 1$. If a_1 is negative, then a branch of super-threshold endemic equilibria exists, and the bifurcation is supercritical. This case is frequently alluded to as a forward bifurcation.

Thus, with the help of Theorem 4.1(iv) of Castillo-Chavez and Song (2004), the following theorem is being concluded:

Theorem 2 *The disease-free equilibrium (DFE) is locally asymptotically stable if R_0 is slightly less than one and if R_0 is slightly greater than one then the DFE is unstable and there is a locally asymptotically stable positive equilibrium near the DFE. Hence the model system (4) exhibits forward bifurcation at $R_0 = 1$.*

The bifurcation from the disease-free equilibrium at $R_0 = 1$ is forward which can be observed from Fig. 2.

Endemic equilibrium and its stability

In this section we investigate the stability of the system (4) at the endemic equilibrium $E_e(S^*, I^*)$.

Equating the second equation of the system (4) to zero, we get:

$$\frac{\beta S^* I^*}{1+\alpha S^*+\gamma I^*} - (\mu+d+\theta)I^* - \frac{aI^{*2}}{bI^{*2}+cI^*+1} = 0 \tag{7}$$

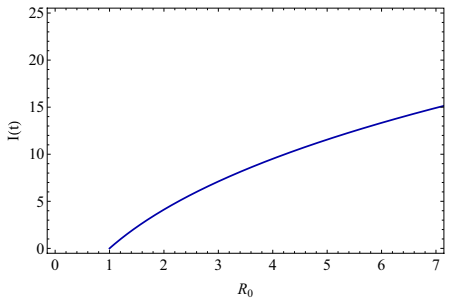


Fig. 2 Figure of $I(t)$ versus R_0. The parameter values are listed in Table 1

After solving the above Eq. (7), we get S^* in terms of I^* as follows:

$$S^* = \frac{(1+\gamma I^*)(p(bI^{*2}+cI^*+1)+aI^*)}{(\beta-p\alpha)(bI^{*2}+1)+I^*(c(\beta-p\alpha)-a\alpha)} \tag{8}$$

where

$p = \mu + d + \theta$

$S^* > 0$ if

$$c(\beta - p\alpha) - \alpha a > 0 \tag{9}$$

This condition also implies that

$$\beta - p\alpha > 0. \tag{10}$$

Now adding first and second equations of the system (4) and equate it to zero, we get :

$$\pi - \mu S^* - (\mu+d+\theta)I^* - \frac{aI^{*2}}{bI^{*2}+cI*+1} = 0 \tag{11}$$

Substituting the value of S^* from Eq. (8) into Eq. (11), we get the following equation in I^* :

$$A_0 + A_1 I^* + A_2 I^{*2} + A_3 I^{*3} + A_4 I^{*4} + A_5 I^{*5} = 0 \tag{12}$$

where

$$\begin{aligned} A_0 &= p(\mu+\pi\alpha)(1-R_0), \\ A_1 &= a(\mu+\pi\alpha) + (\beta-p\alpha)(p-2c\pi) + p\mu(2c+\gamma), \\ A_2 &= ac(\mu+\pi\alpha) + 2pc(\beta-p\alpha+\gamma\mu) + p\mu(2b+c^2) \\ &\quad + a(\beta+\gamma\mu) - (2\alpha p\alpha + \pi(2b+c^2)(\beta-p\alpha)), \\ A_3 &= ab(\mu+\pi\alpha) + pc^2\gamma\mu + 2bp\mu(c+\gamma) \\ &\quad + (\beta-p\alpha)(c(a+pc)+2b(p-c\pi)) - a^2\alpha - ac(p\alpha-\gamma\mu), \\ A_4 &= ab(\beta-2p\alpha+\gamma\mu) + b(\beta-p\alpha)(2pc-b\pi) \\ &\quad + bp\mu(b+2c\gamma), \\ A_5 &= pb^2(\beta-\alpha p+\gamma\mu). \end{aligned}$$

With the help of Descartes' rule of signs (Wang 2004), Eq. (12) has a unique positive real root I^* if any one of the following holds:

(i) $A_5 > 0$, $A_4 < 0$, $A_3 < 0$, $A_2 < 0$, $A_1 < 0$ and $A_0 < 0$.
(ii) $A_5 > 0$, $A_4 > 0$, $A_3 < 0$, $A_2 < 0$, $A_1 < 0$ and $A_0 < 0$.
(iii) $A_5 > 0$, $A_4 > 0$, $A_3 > 0$, $A_2 < 0$, $A_1 < 0$ and $A_0 < 0$.
(iv) $A_5 > 0$, $A_4 > 0$, $A_3 > 0$, $A_2 > 0$, $A_1 < 0$ and $A_0 < 0$.
(v) $A_5 > 0$, $A_4 > 0$, $A_3 > 0$, $A_2 > 0$, $A_1 > 0$ and $A_0 < 0$. (13)

After determining the value of I^*, we can determine the value of S^* from Eq. (8). This implies that there exists a unique positive endemic equilibrium $E_e(S^*, I^*)$ if one of the conditions (13) hold.

If conditions (13) are not satisfied, then we obtain the following result:

Proposition 1 *If $R_0 > 1$, then there is either a unique or three or five positive endemic equilibria, if all equilibria are simple roots.*

Proof Suppose $R_0 > 1$. From Eq. (12) we have fifth degree polynomial in I^*:

$$F(I^*) = A_0 + A_1 I^* + A_2 I^{*2} + A_3 I^{*3} + A_4 I^{*4} + A_5 I^{*5}.$$

Leading coefficient of I^* is $A_5 = pb^2(\beta - \alpha p + \gamma\mu) > 0$. Hence

$$\lim_{I^*\to\infty} F(I^*) = +\infty.$$

Also note that $F(0) = A_0$ and $A_0 < 0$ if $R_0 > 1$. $F(I^*)$ is a continuous function on I^* and by fundamental theorem of algebra, we know that this polynomial can have at most five real roots. □

We now analyze the local stability of endemic equilibrium E_e as follows:

The characteristic equation of the system (4) evaluated at E_e is a second-degree transcendental equation:

$$\lambda^2 + p_0\lambda + q_0 + (p_1\lambda + q_1)e^{-\lambda\tau} = 0 \tag{14}$$

where

$$\begin{aligned} p_0 &= 2\mu + d + \theta + \frac{acI^{*2}+2aI^*}{bI^{*2}+cI^*+1}, \\ q_0 &= \mu\left(\mu + d + \theta + \frac{acI^{*2}+2aI^*}{bI^{*2}+cI^*+1}\right), \\ p_1 &= \frac{\beta}{(1+\alpha S^*+\gamma I^*)^2}(I^*(1+\gamma I^*) - S^*(1+\alpha S^*)), \\ q_1 &= -\frac{\mu\beta S^*(1+\alpha S^*)}{(1+\alpha S^*+\gamma I^*)^2} + \frac{\beta I^*(1+\gamma I^*)}{(1+\alpha S^*+\gamma I^*)^2} \\ &\quad \left(\mu + d + \theta + \frac{acI^{*2}+2aI^*}{bI^{*2}+cI^*+1}\right). \end{aligned}$$

Theorem 3 *At $\tau = 0$, E_e is locally asymptotically stable if $\frac{S^*}{I^*} \leq \frac{1+\gamma I^*}{1+\alpha S^*}$ is satisfied.*

Proof The characteristic equation for the endemic equilibrium E_e at $\tau = 0$ is given by

$$\lambda^2 + p_0\lambda + q_0 + (p_1\lambda + q_1) = 0.$$

It is easy to verify that if $\frac{ac-aI^{*2}}{(I^{*2}+bI^*+c)^2} \geq \frac{\beta S^*(1+\alpha S^*)}{(1+\alpha S^*+\gamma I^*)^2}$ is satisfied, then

$$\begin{aligned} p_0 + p_1 &= 2\mu + d + \theta + \frac{acI^{*2} + 2aI^*}{\left(bI^{*2} + cI^* + 1\right)^2} \\ &\quad + \frac{\beta}{(1+\alpha S^* + \gamma I^*)^2}(I^*(1+\gamma I^*) - S^*(1+\alpha S^*)) > 0, \\ q_0 + q_1 &= \mu(\mu + d + \theta) + \frac{aI^*(2+cI^*)}{\left(bI^{*2} + cI^* + 1\right)^2} \\ &\quad \left(\mu + \frac{\beta I^*(1+\gamma I^*)}{(1+\alpha S^* + \gamma I^*)^2}\right) + \frac{\beta}{(1+\alpha S^* + \gamma I^*)^2} \\ &\quad ((d+\theta)I^*(1+\gamma I^*) + \mu(I^*(1+\gamma I^*) - S^*(1+\alpha S^*))) > 0. \end{aligned}$$

Hence, by Routh–Hurwitz criterion, it is concluded that the endemic equilibrium E_e of the system (4) is locally asymptotically stable when $\tau = 0$. □

Theorem 4 *For $\tau > 0, E_e$ is locally asymptotically stable if the conditions $\mu(1+\alpha S^* + \gamma I^*)^2 \geq \beta I^*(1+\gamma I^*)$ and $2(\mu + d + \theta)I^*(1+\gamma I^*) \geq \mu S^*(1+\alpha S^*)$ are satisfied simultaneously.*

Proof The characteristic equation for $\tau > 0$ evaluated at E_e is given by Eq. (14) which is

$$\lambda^2 + p_0\lambda + q_0 + (p_1\lambda + q_1)e^{-\lambda\tau} = 0.$$

For $\tau > 0$, by applying theorem 2.4 in Ruan and Wei (2003), it follows that the characteristic root of (14) must pass through the imaginary axis for the occurrence of instability for fixed value of time delay τ. Assume that $\lambda = i\eta, \eta > 0$ is the root of the characteristic Eq. (14).

Substituting $\lambda = i\eta$ in Eq. (14), we obtain

$$\left(-\eta^2 + q_0 + p_1\eta\sin\eta\tau + q_1\cos\eta\tau\right) + i\left(p_0\eta + p_1\eta\cos\eta\tau - q_1\sin\eta\tau\right) \tag{15}$$

After separation of real and imaginary parts of Eq. (15), we get

$$p_1\eta\sin\eta\tau + q_1\cos\eta\tau = \eta^2 - q_0, \tag{16}$$

$$p_1\eta\cos\eta\tau - q_1\sin\eta\tau = -\eta p_0. \tag{17}$$

Eliminating τ by squaring and adding equations (16) and (17), we obtain a biquadratic polynomial in η as

$$\eta^4 + \left(p_0^2 - 2q_0 - p_1^2\right)\eta^2 + \left(q_0^2 - q_1^2\right) = 0. \tag{18}$$

Letting $\eta^2 = x$, Eq. (18) becomes:

$$x^2 + Ax + B = 0 \tag{19}$$

where

$$\begin{aligned} A &= p_0^2 - 2q_0 - p_1^2, \quad B = q_0^2 - q_1^2. \\ A &= \left(2\mu + d + \theta + \frac{acI^{*2} + 2aI^*}{bI^{*2} + cI^* + 1}\right)^2 - 2\mu\left(\mu + d + \theta + \frac{acI^{*2} + 2aI^*}{bI^{*2} + cI^* + 1}\right) \\ &\quad - \frac{\beta^2}{(1+\alpha S^* + \gamma I^*)^4}\left(I^*\left(1+\gamma I^*\right) - S^*\left(1+\alpha S^*\right)\right)^2 \\ &= (\mu + d + \theta)^2 + 2(\mu + d + \theta)\frac{acI^{*2} + 2aI^*}{bI^{*2} + cI^* + 1} + \left(\frac{acI^{*2} + 2aI^*}{bI^{*2} + cI^* + 1}\right)^2 + \mu^2 \\ &\quad - \left(\frac{\beta}{(1+\alpha S^* + \gamma I^*)^2}\left(I^*(1+\gamma I^*) - S^*(1+\alpha S^*)\right)\right)^2. \\ B &= q_0^2 - q_1^2 \\ &= \left(\mu\left(\mu + d + \theta + \frac{acI^{*2} + 2aI^*}{bI^{*2} + cI^* + 1}\right)\right)^2 \\ &\quad - \left(-\frac{\mu\beta S^*(1+\alpha S^*)}{(1+\alpha S^* + \gamma I^*)^2} + \frac{\beta I^*(1+\gamma I^*)}{(1+\alpha S^* + \gamma I^*)^2}\left(\mu + d + \theta + \frac{acI^{*2} + 2aI^*}{bI^{*2} + cI^* + 1}\right)\right)^2 \\ &= \left(\mu + d + \theta + \frac{acI^{*2} + 2aI^*}{bI^{*2} + cI^* + 1}\right)^2\left(\mu^2 - \frac{\beta^2 I^{*2}(1+\gamma I^*)^2}{(1+\alpha S^* + \gamma I^*)^4}\right) \\ &\quad + \frac{2\mu\beta^2 S^*(1+\alpha S^*)I^*(1+\gamma I^*)(acI^{*2} + 2aI^*)}{(1+\alpha S^* + \gamma I^*)^4(bI^{*2} + cI^* + 1)} \\ &\quad + \frac{\mu\beta^2 S^*(1+\alpha S^*)}{(1+\alpha S^* + \gamma I^*)^4}\left(2(\mu + d + \theta)I^*(1+\gamma I^*) - \mu S^*(1+\alpha S^*)\right). \end{aligned}$$

$A > 0$ if and only if $\mu(1+\alpha S^* + \gamma I^*)^2 \geq \beta(I^*(1+\gamma I^*) - S^*(1+\alpha S^*))$ and $B > 0$ if and only if both the conditions $\mu(1+\alpha S^* + \gamma I^*)^2 \geq \beta I^*(1+\gamma I^*)$ and $2(\mu + d + \theta)I^*(1+\gamma I^*) \geq \mu S^*(1+\alpha S^*)$ are satisfied simultaneously.

Note that $\mu(1+\alpha S^* + \gamma I^*)^2 \geq \beta I^*(1+\gamma I^*)$ implies that:

$$\mu(1+\alpha S^* + \gamma I^*)^2 \geq \beta(I^*(1+\gamma I^*) - S^*(1+\alpha S^*)).$$

According to Routh–Hurwitz criterion, a contradiction arises with the assumption of instability i.e $\lambda = i\eta$. Thus, it can be concluded that the endemic equilibrium E_e of the system (4) is locally asymptotically stable for $\tau > 0$. □

Hopf bifurcation analysis

If $B = q_0^2 - q_1^2$ in Eq. (19) is negative, then there is a unique positive η_0 satisfying Eq. (19), i.e., there is single pair of purely imaginary roots $\pm i\eta_0$ to Eq. (14).

From Eqs. (16) and (17), τ_n corresponding to η_0 can be obtained as

$$\tau_n = \frac{1}{\eta_0}\arccos\left(\frac{\eta_0^2(q_1 - p_0p_1) - q_0q_1}{p_1^2\eta_0^2 + q_1^2}\right) + \frac{2n\pi}{\eta_0}, \quad n = 0, 1, 2\ldots \tag{20}$$

Endemic equilibrium E_e is stable for $\tau < \tau_0$ if transversality condition holds, i.e., $\frac{d}{d\tau}(Re\lambda)\Big|_{\lambda=i\eta_0} \neq 0$.

Differentiating Eq. (14) with respect to τ, we get

$$\begin{aligned} &\left(2\lambda + p_0 + p_1e^{-\lambda\tau} - (p_1\lambda + q_1)\tau e^{-\lambda\tau}\right)\frac{d\lambda}{d\tau} = \lambda(p_1\lambda + q_1)e^{-\lambda\tau} \\ &\left[\frac{d\lambda}{d\tau}\right]^{-1} = \frac{2\lambda + p_0 + p_1e^{-\lambda\tau} - (p_1\lambda + q_1)\tau e^{-\lambda\tau}}{\lambda(p_1\lambda + q_1)e^{-\lambda\tau}} \\ &= \frac{2\lambda + p_0}{\lambda(p_1\lambda + q_1)e^{-\lambda\tau}} + \frac{p_1}{\lambda(p_1\lambda + q_1)} - \frac{\tau}{\lambda} \\ &= \frac{2\lambda + p_0}{-\lambda(\lambda^2 + p_0\lambda + q_0)} + \frac{p_1}{\lambda(p_1\lambda + q_1)} - \frac{\tau}{\lambda}. \end{aligned} \tag{21}$$

$$\begin{aligned} &\frac{d}{d\tau}(Re\lambda)\Big|_{\lambda=i\eta_0} = Re\left(\frac{d\lambda}{d\tau}\right)^{-1}\Big|_{\lambda=i\eta_0} \\ &= Re\left(\frac{2i\eta_0 + p_0}{-i\eta_0(-\eta_0^2 + ip_0\eta_0 + q_0)} + \frac{p_1}{-p_1\eta_0{}^2 + iq_1\eta_0} + \frac{i\tau}{\eta_0}\right) \\ &= \frac{1}{\eta_0}\left(\frac{2\eta_0(\eta_0^2 - q_0) + p_0^2\eta_0}{(p_0\eta_0)^2 + (\eta_0^2 - q_0)^2} - \frac{p_1^2\eta_0}{(p_1\eta_0)^2 + q_1^2}\right) \\ &= \frac{2(\eta_0^2 - q_0) + p_0^2}{(p_0\eta_0)^2 + (\eta_0^2 - q_0)^2} - \frac{p_1^2}{(p_1\eta_0)^2 + q_1^2}. \end{aligned} \tag{22}$$

Now, on squaring and adding Eqs. (16) and (17), we get

$$(p_1\eta_0)^2 + q_1^2 = (p_0\eta_0)^2 + (\eta_0^2 - q_0)^2$$

so that Eq. (22) can be written as

$$\frac{d}{d\tau}(Re\lambda)\Big|_{\lambda=i\eta_0} = \frac{2\eta_0^2 + (p_0^2 - 2q_0 - p_1^2)}{p_1^2\eta_0^2 + q_1^2}. \tag{23}$$

Under the condition $A = p_0^2 - 2q_0 - p_1^2 > 0$, it can be seen that

$$\frac{d}{d\tau}(Re\lambda)\Big|_{\lambda=i\eta_0} > 0.$$

Therefore, the transversality condition holds and Hopf bifurcation occurs at $\eta = \eta_0, \tau = \tau_0$.

Summarizing the above analysis, we arrive at the following theorem:

Theorem 5 *If condition $q_0^2 - q_1^2 < 0$ holds, the endemic equilibrium E_e of the system (4) is asymptotically stable for $\tau \in [0, \tau_0)$ and it undergoes Hopf bifurcation at $\tau = \tau_0$.*

Numerical simulation

Since it is important to analyze the dynamical behavior of the model, in this section, the system (4) is integrated numerically using the set of tested parameters given in Table 1.

The computer simulations are performed for S and I for various values of τ. The trajectory of S and I with initial conditions $S(0) = 33, I(0) = 5$ approaches the endemic equilibrium as shown in Fig. 3.

Figure 3 shows the effect of time delay on susceptible population and infective population, respectively, with respect to time for the set of parameters given in Table 1 for different values of time delay τ. It is shown that as the time delay τ increases, the number of susceptible starts

Table 1 List of parameters

Parameter	Interpretation	Value	References
π	Recruitment rate of susceptible	2	Dubey et al. (2016)
α	Measure of inhibition taken by susceptibles	0.004	Dubey et al. (2016)
β	Transmission rate	0.00924	Assumed
μ	Natural death rate	0.05	Dubey et al. (2016)
d	Disease-induced death rate	0.001	Dubey et al. (2016)
γ	Measure of inhibition taken by infectives	0.002	Dubey et al. (2016)
θ	Recovery rate	0.002	Dubey et al. (2016)
a	Treatment rate of infected individuals/cure rate	0.002	Assumed
b	Limitation rate in treatment availability	0.0005	Assumed
c	Saturation constant	0.0002	Assumed

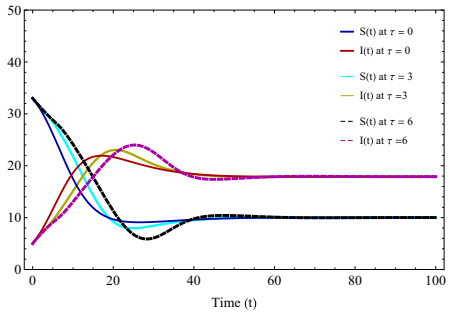


Fig. 3 Susceptible and infected population for various values of τ

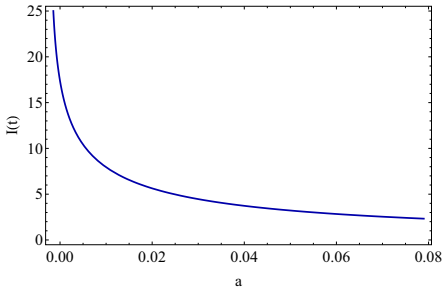


Fig. 5 Infectives $I(t)$ versus treatment rate a

decreasing and number of infectives starts growing. Due to interplay between the number of infectives and susceptibles, infectives settle to its steady state but never approaches to zero, which shows that system approaches endemic equilibrium $E_e(9.8292, 18.4177)$.

Figure 4 shows the influence of transmission rate β on infected population for $\tau = 1$ day. It is true that the higher the effective contact rate, the higher will be the possibility of spreading the disease. In Fig. 4 we note that when the effective contact rate β is high, then more people will be infected, and when the effective contact rate β is low, then fewer people are infected.

Figure 5 depicts the behavior of infected population $I(t)$ with respect to cure rate a. From this figure, it can be seen that infected population is decreasing with the increment in treatment rate a.

Figure 6 shows the effect of measure of inhibitions taken by susceptible and infectives respectively. These figures show that when the inhibition is less, then more number of people are getting infected, and when inhibition is more, then fewer people are getting infected.

Figure 7 shows the effect of saturated treatment rate on the infected individuals for the time lag $\tau = 1$ day. The treatment is an imperative strategy to diminish the spread of diseases. This figure shows that saturated treatment rate is diminishing the infection.

In Fig. 8a, a plot is drawn for infected and susceptible population versus time t for time lag $\tau = 12$ days. Figure 8b shows the phase plot between susceptible $S(t)$ and infected $I(t)$ population which shows the limit cycle for time lag $\tau = 12$ days. Using parameter values given in Table 1, it clear from Fig. 8a, b that endemic equilibrium $E_e = (9.8292, 18.4177)$ is the unique endemic equilibrium. According to the algorithm given in Eq. (20) and Theorem 5, we can compute $\tau_0 = 12.5801$, $p_0^2 - 2q_0 - p_1^2 = 0.0109139 > 0$ and $q_0^2 - q_1^2 = -0.000150942 < 0$. Clearly, from Fig. 8 it can be seen that when $\tau = 12 < \tau_0 = 12.5801$, then the endemic equilibrium is asymptotically stable.

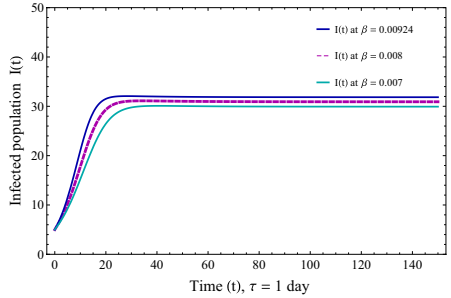
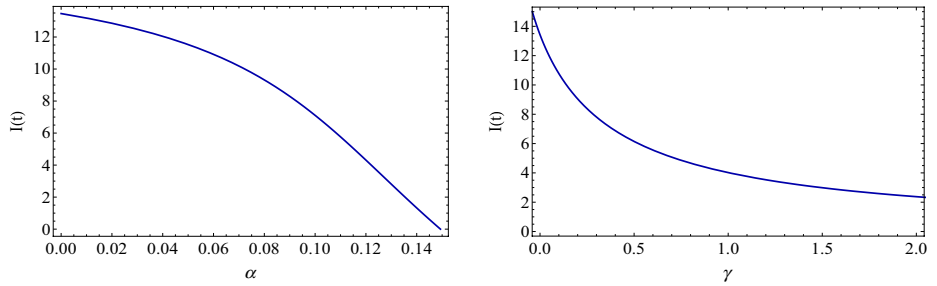


Fig. 4 Behavior of infected population for various values of β at $\tau = 1$ day

Conclusion

In conclusion, this paper has, as the fundamental objective, the formulation of a nonlinear SIR mathematical model to study the role of Beddington–DeAngelis-type nonlinear incidence rate, by incorporating a delay time to analyze the equilibrium points and their stability and saturated treatment rate-type nonlinear treatment rate. It is assumed that there is a time lag due to latency period of pathogens, i.e., the development of an infection from the time the pathogen enters the body until symptoms first appear. The model exhibits two equilibria; the disease-free and endemic equilibrium. The local stability of the disease-free equilibrium is determined by the basic reproduction number R_0. The disease-free equilibrium has been shown to be stable for $R_0 < 1$, i.e., disease dies out for $R_0 < 1$ and for $R_0 > 1$, it becomes unstable and



a Infectives $I(t)$ versus Inhibition effect α. **b** Infectives $I(t)$ versus Inhibition effect γ.

Fig. 6 Effects of measure of inhibitions

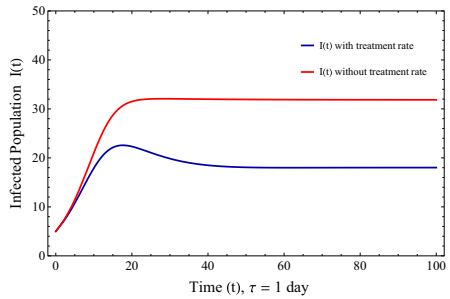
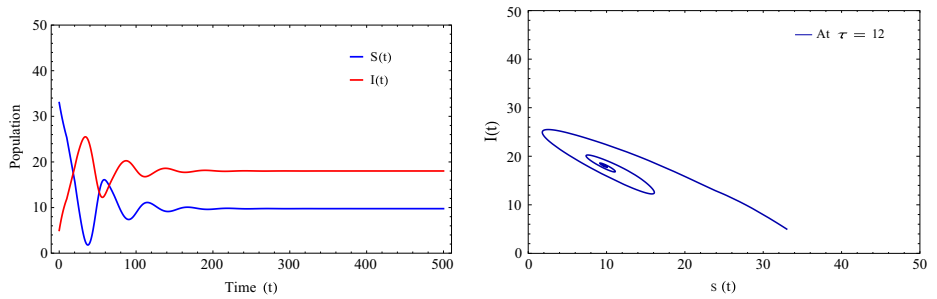


Fig. 7 Infected population with and without saturated treatment rate at $\tau = 1$ day

the endemic equilibrium exists. We also discuss the stability of disease-free equilibrium at $R_0 = 1$ using center manifold theory. We observed that at $R_0 = 1$, model exhibits forward bifurcation and changes its stability as R_0 crosses one. The stability analysis demonstrates that endemic equilibrium is locally asymptotically stable under certain conditions for the time lag $\tau \geq 0$ as stated in Theorems 3 and 4. Further, system (4) has been simulated numerically for the effect of time delay τ and showed that as delay increases, infected population also increases. Furthermore, we simulated the model numerically to see the effects of transmission rate, measure of inhibition and treatment rate. From the graphs, it is depicted that the infection can be eradicated from the society if the treatment given to the population is managed according to saturated treatment rate. Also, analytical and numerical results showed that oscillatory behavior of the infected population will also occur, indicating the existence of a Hopf bifurcation.



a Plot of $S(t)$ and $I(t)$ with respect to time t. **b** $I(t)$ versus $S(t)$.

Fig. 8 Behavior of susceptible and infected population for the time lag $\tau = 12$ days

Acknowledgements The authors acknowledged the support of Delhi Technological University, Delhi, India, for giving monetary help to complete this research work. They are also indebted to the anonymous reviewers and the handling editor for their constructive comments and suggestions which have enhanced the paper.

References

Anderson RM, May RM (1978) Regulation and stability of host-parasite population. Interactions: I. Regulatory processes. J Anim Ecol 47:219–267
Anderson RM, May RM (1992) Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford
Bailey NTJ (1975) The mathematical theory of infectious diseases and its applications. Griffin, London
Beddington JR (1975) Mutual interference between parasites or predators and its effect on searching efficiency. J Anim Ecol 44:331–340
Brauer F, Castillo-Chavez C (2001) Mathematical models in population biology and epidemiology. Springer, New York
Capasso V, Serio G (1978) A generalization of the Kermack–Mckendrick deterministic epidemic model. Math Biosci 42(1–2):43–61
Castillo-Chavez C, Song B (2004) Dynamical models of tuberculosis and their applications. Math Biosci Eng 1(2):361–404
DeAngelis DL, Goldstein RA, O'Neill RV (1975) A model for tropic interaction. Ecology 56:881–892
Dubey B, Patra A, Srivastava PK, Dubey US (2013) Modeling and analysis of an SEIR model with different types of nonlinear treatment rates. J Biol Syst 21(03):1350023
Dubey B, Dubey P, Dubey US (2015) Dynamics of an SIR model with nonlinear incidence and treatment rate. Appl Appl Math 10(2):718–737
Dubey P, Dubey B, Dubey US (2016) An SIR model with nonlinear incidence rate and Holling type III treatment rate. In: Cushing J, Saleem M, Srivastava H, Khan M, Merajuddin M (eds) Applied analysis in biological and physical sciences. Springer proceedings in mathematics and statistics, vol 186. Springer, New Delhi, pp 63–81
Elaiw AM, Azoz SA (2013) Global properties of a class of HIV infection models with Beddington–DeAngelis functional response. Math Methods Appl Sci 36:383–394
Gumel AB, McCluskey CC, Watmough J (2007) An SVEIR model for assessing potential impact of an imperfect anti-SARS vaccine. Math Biosci Eng 3(3):485–512
Hattaf K, Lashari AA, Louartassi Y, Yousfi N (2013) A delayed SIR epidemic model with general incidence rate. Electron J Qual Theory Differ Equ 3:1–9
Hethcote HW, van den Driessche P (1995) An SIS epidemic model with variable population size and a delay. J Math Biol 34(2):177–194
Kaddar A (2010) Stability analysis in a delayed SIR epidemic model with a saturated incidence rate. Nonlinear Anal Model Control 15:299–306
Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. Proc R Soc Lond A 115:700–721
Korobeinikov A (2007) Global properties of infectious disease models with nonlinear incidence. Bull Math Biol 69(6):1871–1886
Korobeinikov A, Maini PK (2005) Nonlinear incidence and stability of infectious disease models. Math Med Biol 22:113–128
Kuang Y (1993) Delay differential equations with applications in population dynamics. Academic Press, Boston
Kumar A, Nilam (2018a) Stability of a time delayed SIR epidemic model along with nonlinear incidence rate and Holling type-II treatment rate. Int J Comput Methods 15(6):1850055
Kumar A, Nilam (2018b) Dynamical model of epidemic along with time delay: Holling type II incidence rate and Monod–Haldane type treatment rate. Differ Equ Dyn Syst. https://doi.org/10.1007/s12591-018-0424-8
Li MY, Muldowney JS (1995) Global stability for the SEIR model in epidemiology. Math Biosci 125:155–164
Li X, Li W, Ghosh M (2009) Stability and bifurcation of an SIR epidemic model with nonlinear incidence and treatment. Appl Math Comput 210:141–150
McCluskey CC (2010) Global stability for an SIR epidemic model with delay and nonlinear incidence. Nonlinear Anal RWA 11(4):3106–3109
Mena-Lorca J, Hethcote HW (1992) Dynamic models of infectious disease as regulators of population size. J Math Biol 30(7):693–716
Mukherjee D (1996) Stability analysis of an S-I epidemic model with time delay. Math Comput Model 24(9):63–68
Ruan S, Wei J (2003) On the zeros of transcendental functions with applications to stability of delay differential equations with two delays. Dyn Contin Discrete Impuls Syst Ser A 10:863–874
Sastry S (1999) Analysis, stability and control. Springer, New York
Song X, Cheng S (2005) A delay-differential equation model of HIV infection of CD4+ T-cells. J Korean Math Soc 42(5):1071–1086
Tipsri S, Chinviriyasit W (2014) Stability analysis of SEIR model with saturated incidence and time delay. Int J Appl Phys Math 4(1):42–45
Van den Driessche P, Watmough J (2002) Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math Biosci 180:29–48
Wang X (2004) A simple proof of Descartes's rule of signs. Am Math Mon. https://doi.org/10.2307/4145072
Wang W (2006) Backward bifurcation of an epidemic model with treatment. Math Biosci 201(1):58–71 pmid:16466756
Wang W, Ruan S (2004) Bifurcations in an epidemic model with constant removal rate of the infectives. J Math Anal Appl 291(2):775–793
Wei C, Chen L (2008) A delayed epidemic model with pulse vaccination. Discrete Dyn Nat Soc. https://doi.org/10.1155/2008/746951
Xu R, Ma Z (2009a) Global stability of a SIR epidemic model with nonlinear incidence rate and time delay. Nonlinear Anal RWA 10(5):3175–3189
Xu R, Ma Z (2009b) Stability of a delayed SIRS epidemic model with a nonlinear incidence rate. Chaos Solitons Fractals 41(5):2319–2325
Yang M, Sun F (2015) Global stability of SIR models with nonlinear Incidence and discontinuous treatment. Electron J Differ Equ 2015(304):1–8
Zhang JZ, Jin Z, Liu QX, Zhang ZY (2008) Analysis of a delayed SIR model with nonlinear incidence rate. Discrete Dyn Nat Soc. https://doi.org/10.1155/2008/636153
Zhang Z, Suo S (2010) Qualitative analysis of an SIR epidemic model with saturated treatment rate. J Appl Math Comput 34:177–194

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

Wireless Personal Communications
https://doi.org/10.1007/s11277-019-06121-7

Adaptive Technique with Cross Correlation for Lowering Signal-to-Noise Ratio Wall in Sensor Networks

Rajiv Kapoor[1] · Rashmi Gupta[2] · Le Hoang Son[3,4] · Sudan Jha[5] · Raghvendra Kumar[6,7]

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Energy detection is considered as most popular spectrum sensing technique. However, the impact of noise uncertainty in wireless environment limits the sensitivity of energy-based detector. It generally suffers from Signal-to-Noise Ratio (SNR) wall, which is defined as the minimum SNR in which robust detection is not possible. This paper presents an adaptive technique to improve SNR wall in sensor networks that combines cross-correlation scheme and dynamic threshold method to improve performance. An analytical expression of SNR wall is derived to show improved performance over traditional energy detector. Theoretical analyses and simulations validate effectiveness of the proposed method under noise uncertainty environment.

Keywords Adaptive threshold · Cognitive radio · Cross-correlation · Power recognition · Noise uncertainty parameters · Signal-to-noise ratio wall

✉ Le Hoang Son
lehoangson@tdtu.edu.vn

Rajiv Kapoor
Rajivkapoor.dtu@gmail.com

Rashmi Gupta
rashmig71@yahoo.com

Sudan Jha
jhasudan@hotmail.com

Raghvendra Kumar
raghvendraagrawal7@gmail.com

1 Department of Electronics & Communication Engineering, Delhi Technological University, Delhi, India

2 Department of Electronics & Communication Engineering, AIACT&R, Delhi, India

3 Division of Data Science, Ton Duc Thang University, Ho Chi Minh City, Vietnam

4 Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam

5 School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, India

6 Computer Science and Engineering Department, LNCT College, Bhopal, MP, India

7 VNU Information Technology Institute, Vietnam National University, Hanoi, Vietnam

Published online: 29 January 2019

1 Introduction

Power recognition can be considered as popular spectrum sensing technique, though the impact of noise uncertainty in wireless environment limits the sensitivity of energy-based detector and generally suffers from the Signal-to-noise ratio (SNR) walls defined as the smallest SNR, where robust detection is not possible [12, 23–25]. The essential benefit of cognitive radio (CR) is its potential toward applying certified frequency spectrum of the primary users (PUs) [31, 55]. CR has emerged as one of the solutions to enhance spectrum performance [11]. It is mainly consisting of two main roles specifically sensing and throughput optimization. Spectrum sensing is improved by enforcing Peak vs. Average Power Ratio of the acquired signal in strength detection [1]. It has some additional advantages such as prior authentication is not required and loosed from sound ambiguity [49]. CR is not effectively relevant to traditional networks. It has the capacity to enhance the overall performance of wireless sensor networks [6, 10].

A cognitive radio detects unused spectrum band and utilizes it in opportunistic way. Energy detection is the most commonly used method to detect presence of unused band [10]. Signal detection at noise uncertainty situation imposes restriction on sensing called the Signal-to-Noise Ratio (SNR) wall [4, 7, 18]. Various methods were proposed for solving the Signal-to-Noise Ratio (SNR) wall problem [14, 30]. Some typical ones include the oversampling with over sampled alerts, the blind spectrum sensing scenario, the non-linear stochastic resonance, adaptive threshold strategies through the most likelihood estimator or second and fourth order second estimator, and correlation strategies [19, 26, 31, 37, 38].

In this article, we proposed a hybrid method to improve the SNR wall. In the first stage, cross-correlation method is used, and in the second stage the adaptive threshold method is used to improve the performance. The main contribution of this article is to derive a new expression for SNR wall by combining the cross-correlation method and adaptive threshold method. We will analyze the shortcomings of the previous works and the idea of innovation in Sect. 2.

The rest of the articles are prepared into four different sections. Section 2 describes related researches for the SNR Wall. Section 3 affords the formula and new model. Section 4 suggests performance analyses and simulation. Finally, conclusions are made in Sect. 5.

2 Related Works

Various methods had been proposed inside the literature for solving the Signal-to-Noise Ratio (SNR) wall problem. Oversampling with over sampled alerts has high correlation among their successive samples compared to noises changed. Later, Shaikh et al. [37] proposed a blind spectrum sensing situation facilitate. Dutta and Manna [9] proposed a new hybrid frame structure for spectrum sensing which has better overall performance than traditional approaches in throughput [10]. The chance of detection at distinct alarm probabilities was evaluated.

Gunes et al. [15] tested resonator parameters, and Liu et al. [27] investigated phenomenon of stochastic resonance in a mono-threshold-system-primarily based detector with additive heritage noise and multiplicative external noise. Ali et al. [2] described an adaptive threshold method with unfixed and adaptive thresholds toward channel situations to

improve detection performance. It calculates a required threshold as a linear feature of recorded sign's general deviation [5]. Zulfikar et al. [55] used cognitive radio generation to handle the problem of inefficiently gadget radio spectrum and boom array use of spectrum. Detection overall performance of energy is relatively structured upon noises even as SNR is very low for PU.

Ali et al. [1] supplied a comprehensive survey on the cognitive radio community paradigm elegant grid, consisting of the tool structural design, statement community composition, application, as well as cognitive radio-based totally communication technology. Possible programs of cognitive radio-primarily based system and cognitive radio-based totally spectrum sensing tactics with their important classifications have been examined with routing and MAC protocols. Some open research challenges faced by cognitive radio-primarily based smart grid networks along have been highlighted therein.

Another technique based totally on adaptive threshold changed into was shown in [19] with spectrum sensing errors being minimized underneath constraints. This comes from the truth that practical systems have spectrum utilization adjustments with wide variety in energetic transmission with corresponding parameters. Oude et al. [31] offered a correlation technique with energy detection. Besides those works, there are a few relevant works on SNR in [11, 20, 36]; Haijun et al. [17].

Nevertheless, they did not reduce SNR Wall as expected. Spectrum sensing schemes are commonly used in many networks. In this paper, the spectrum sensing schemes are used to deal with noise uncertainty and signal-to-noise ratio (SNR) wall. Until the sensing time is not evaluated, we cannot achieve desired performance through signal detection. Therefore, spectrum sensing schemes are used to explore and improve the spectrum's efficiency. They are used for many purposes like identifying frequency holes, assigning frequencies to the spectrum for analyzing of zero harmful interference to the channel, variations in the noise as a function of real time, and enhancing the low signal-to-noise ratio (SNR).

We will show in the Sect. 3 that without the sensor detector, it is unable to achieve the desired performance. Due to this, the robustness against power noise uncertainty will be unpredictably low. Therefore, this scheme is used for both theoretical and experimental analysis to confirm the SNR wall existence under the noise power uncertainty conditions. The spectrum sensing technique used in this article analyzes performance of strength detector that maybe degraded. The noise uncertainty is indeed integrated into the system. In the cross-correlation method, same signs pass into two unique paths which lessen uncorrelated noises. Therefore, SNRs and SNR wall are progressed at outputs. The new hybrid technique based on cross-correlation and adaptive threshold technique effectively treats the drawbacks of existing works. An expression for SNR wall turned into derived that involve to correlation ε, uncertainty U and the dynamic threshold issue d. Simulations will be done to validate the new derived SNR wall in improving energy detection performance in comparison to correlation based methods.

3 Proposed Method

3.1 Problem Formulation

Spectrum sensing using power recognition can be considered as the binary hypothesis testing problem given as:

(i) H_0: represent that the primary user is absent.
(ii) H_1: represent that the primary user is present.

Let $S(N)$ be transmitted primary user through channel $l$and $h = 0$. The received signal $Y(N)$ is given by,

$$Y(N) = S(N) + W(N), \tag{1}$$

where $W(N)$ be an additive white Gaussian noise. Let $D(Y)$ denotes the decision statistics to decide H_1 or H_0.

$$\begin{aligned} &1D(Y) = \frac{1}{N}\sum_{n=0}^{n} \mathrm{Y(N)} \\ &D(Y) \geq \gamma \ : H_1 \\ &D(Y) < \gamma \ : H_0 \end{aligned} \tag{2}$$

Noise variance is completely known. From the middle limit theorem, the subsequent approximation is preserved [36].

$$D\left(Y/H_0\right) \sim N\left(\sigma_n^2, 2\sigma_n^4/N\right) \tag{3}$$

$$D\left(Y/H_1\right) \sim N\left(p + \sigma_n^2, (p + 2\sigma_n^2)^2/N\right) \tag{4}$$

Different probabilities terms such as possibility of discovery $P_d = \mathrm{prob}\left\{\mathrm{Decision} = H_1/H_1\right\}$, false alarm probability $P_f = \mathrm{prob}\left\{\mathrm{Decision} = H_1/H_0\right\}$ and probability of miss $P_m = \mathrm{prob}\left\{\mathrm{Decision} = H_0/H_1\right\}$ are given in Eqs. (5–6).

$$P_{FA} = P_r\left(D(Y/H_0) > \gamma\right) = Q\left(\frac{\gamma - \sigma_n^2)}{\sqrt{2\sigma_n^4/N}}\right) \tag{5}$$

$$P_D = Q\left(\frac{\gamma - (p + \sigma_n^2)}{\sqrt{2(p + \sigma_n^2)^2/N}}\right) \tag{6}$$

From Eqs. (5) and (6), expression for N can be calculated to achieve desired P_d, by fixing P_{FA} to a certain value and for known SNR value. Quantity of sample essentially is given by Eq. (7).

$$N = 2\left[Q^{-1}(P_{FA}) - Q^{-1}(P_D)\right]^2 \times SNR^{-2} \tag{7}$$

However, the noise influences are predictable with distributional ambiguity in a particular gap:

$$\sigma_n^2 \in \left[\frac{\sigma_n^2}{\mu}, \mu\sigma_n^2\right]$$

where $\mu > 1$ is the noise uncertainty coefficient and $U = 10\log_{10}\mu$. From (7), we obtain

$$\begin{aligned} &N = 2\left[\mu Q^{-1}(P_{FA}) - \left(\frac{1}{\mu} + SNR\right)Q^{-1}(P_D)\right]^2 \times \left(SNR - \left(\mu - \frac{1}{\mu}\right)\right)^{-2} \\ &N \to \infty, SNR \to \frac{\mu^2 - 1}{\mu^2} \end{aligned} \tag{8}$$

The **SNR Wall** in terms noise uncertainty factor is given as,

$$SNR_{Wall} = \frac{\mu^2 - 1}{\mu^2} \tag{9}$$

It is recognized that signal detection SNR_{Wall} is unreliable.

3.2 Proposed Hybrid Method

Figure 1 shows block diagram for the proposed method in which both cross-correlation and dynamic threshold techniques are employed simultaneously to improve SNR Wall. Let us transmit PU signal $S(t)$ distorted with channel noise $n(t)$. Here, $n(t) = n_c(t) + n_I(t)$ where $n_c(t)$ is correlative noise component and $n_I(t)$ is interference noise component ($i = 1$, 2-signal travelling through path 1 and path 2).

The *first part* of the block diagram shows the cross-correlation method, as signal travels through path1 and path2 of receiver cross-correlation will reduce interference noise. The *second part* of the method shows the adaptive threshold based on energy detector which further improves SNR and detects unused primary user signals. Let ε be the noise correlation between two receivers.

Let us denote,

$$\begin{aligned} &psd \text{ of } n_c(t) = \varepsilon N_0 \\ &psd \text{ of } n_1(t) = psd \text{ of } n_2(t) = (1 - \varepsilon)N_0 \end{aligned} \tag{10}$$

At the output, SNR $= P_s/N_o W.N_o W$ is the noise within band of interest. From Fig. 1, we clearly see that

$$\mathrm{Y} = Y_{\mathrm{re}} + Y_{\mathrm{im}} \tag{11}$$

where Y_{re} and Y_{im} are the real and imaginary parts of the sample respectively because every samples are independent and identically distributed (IID), after that the central limit theorem for a large number of samples, their common converge in the direction of a standard allocation. Here, we are assuming that both receivers in Fig. 1 are perfectly in phase i.e. $\Delta\emptyset = 0$. Therefore, imaginary parts can be assumed as zero. For simplicity purpose, we are taking $N_o = 1$.

Case 1: Under hypothesis H_0 (PU absent $SNR = 0$), mean and variance are given as,

$$\mu_0 = \varepsilon \text{ and } \sigma_0^2 = \frac{1 + \varepsilon^2}{2N} \tag{12}$$

Case 2: Under hypothesis H_1 (PU is present $SNR > 0$), mean and variance are given as

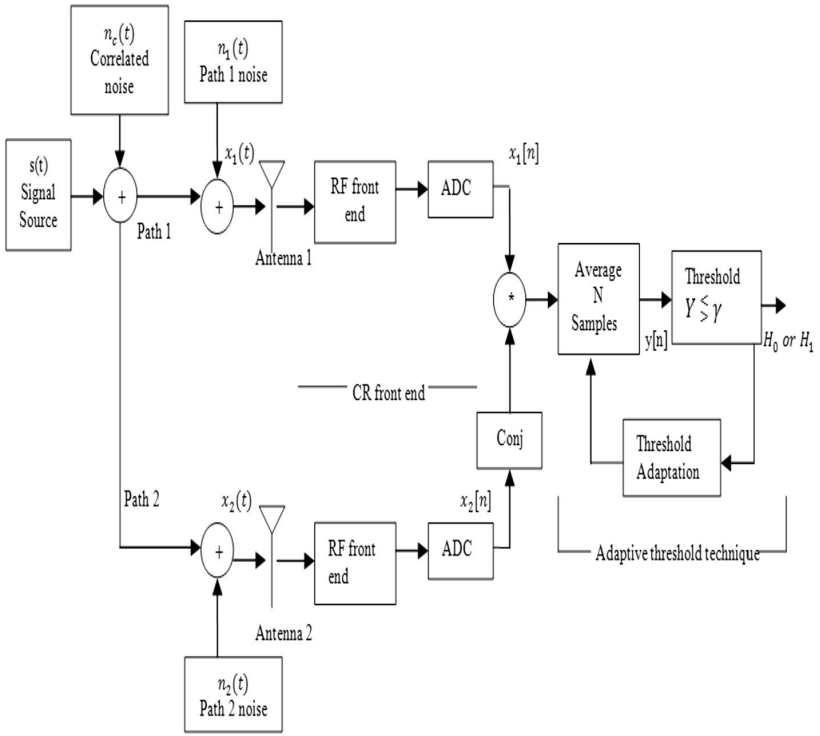


Fig. 1 Architecture of the proposed method

$$\mu_1 = \varepsilon + SNR$$
$$\sigma_1^2 = \frac{2SNR^2 + (2+2\varepsilon)SNR + 1 + \varepsilon^2}{2N} \tag{13}$$

P_{FA} and P_D for fixed threshold energy detector method are given as,

$$P_{FA} = Q\left(\frac{\gamma - \mu_0}{\sigma_0}\right) \tag{14}$$

$$P_D = Q\left(\frac{\gamma - \mu_1}{\sigma_1}\right) \tag{15}$$

Due to noise uncertainty for the fixed threshold method, performance of detector also degrades. Assume threshold γ is varied in such a way that system performance is maintained constant to its previous desired value, Here, d is called as dynamic threshold factor ($d > 1$) then (14) and (15) are modified as,

$$P_{FA} = Q\left(\frac{\gamma' - \mu_0}{\sigma_0}\right) = Q\left(\frac{\gamma d - \mu_0}{\sigma_0}\right) \tag{16}$$

$$P_D = Q\left(\frac{\gamma' - \mu_1}{\sigma_1}\right) = Q\left(\frac{\gamma'/d - \mu_1}{\sigma_1}\right) \tag{17}$$

From Eq. (17), for desired value of P_{FA}, setting the value of γ and multiplying with U factor for the evaluation under the noise uncertainty condition is as below:

$$\gamma = U\frac{[\mu_0 + \sigma_0 Q^{-1}(P_{FA,des})]}{d} \tag{18}$$

Putting value of γ in Eq. (17), we obtain

$$P_D = Q\left(\frac{\frac{[\mu_0 + \sigma_0 Q^{-1}(P_{FA,des})]}{d^2} - \mu_1}{\sigma_1}\right) \tag{19}$$

Putting values of μ_0, μ_1, σ_0 , σ_1 in Eq. (19), we gain

$$P_D \approx Q\left[\frac{\left(\varepsilon\left(U - d^2\right) - d^2 SNR\right)\sqrt{N} + U\sqrt{P}Q^{-1}\left(P_{FA}, des\right)}{d^2\sqrt{\gamma + (\varepsilon+1)SNR}}\right] \tag{20}$$
$$p \equiv 0.5(1+\varepsilon^2),\ SNR \ll 1 \& N \gg 1$$

For $SNR_{\min}$ expression, we approximate Eq. (20) for $N \gg 1$ and given P_D and P_{FA}.

$$SNR_{\min} \approx \frac{\varepsilon\left(U - d^2\right)}{d^2} + \frac{X^2(\varepsilon+1)}{2N} + \frac{UpY}{d^2} - \frac{X\sqrt{Nd^6\left(pd^2 + \left(U + d^2\right)\left(\varepsilon + \varepsilon^2\right)\right) + Z}}{N}$$
$$(P_D \geq P_{D,des}) \,\&\, (P_{FA} \leq P_{FA,des})$$
$$X \equiv Q^{-1}(P_{D,des}),\ Y \equiv Q^{-1}(P_{FA,des})$$
$$Z \equiv X^2(\varepsilon+1)^2 d^8 + pUY(\varepsilon+1)\sqrt{N}d^6 \tag{21}$$

The above expression can be simplified by using ($N \gg 1$) and $|X|$ and $|Y|$ are fairly small such that ($|X| \cdot |Y| \leq 10$), which gives $P_{FA} = P_{MD} \geq 10^{-23}$.

$$SNR_{\min} \approx \frac{\varepsilon\left(U - d^2\right)}{d^2} + \frac{UpY - Xd^5\sqrt{pd^2 + \varepsilon\left(U - d^2\right)(1+\varepsilon)}}{d^2\sqrt{N}} \tag{22}$$

Since SNR Wall is defined as a minimum SNR for infinitely long averaging below which a signal cannot be detect. Thus, SNR Wall is expressed as,

$$SNR_{wall_new} = \log_{N\to\infty}\left(SNR_{min}\right) = \frac{\varepsilon\left(U - d^2\right)}{d^2} \tag{23}$$

For the fixed threshold model ($d = 1$), the above expression becomes

$$snr_{wall_new} = \varepsilon(U - 1) \tag{24}$$

The mathematical expression given in Eq. (23) represents the SNR Wall for proposed method which depends upon three factors:

1. Uncertainty factor (U).
2. Cross-correlation between two paths (ε).
3. Dynamic threshold factor (d).

As the value of U increases the SNR Wall, to compensate the effect of U, the value of ε should maintain low and the value of d should be greater than 1. SNR Wall for $d = 1$, shown in Eq. (24) which represents for fixed threshold energy detection method.

4 Experimental Results

This section describes performance of proposed algorithm to verify the accuracy of the derived relation on SNR degradation and packet loss charge. The SNR_{wall} are verified by the real scenario simulation performed using MATLAB in accordance with the parameters presented in IEEE 802.22 standards. Results obtained from previous techniques [26, 28] are compared with those of the proposed method. Assuming value of P_{MD} and $P_{FA} = 10^{-5}$, performance of SNR_{min} is plotted in Fig. 2 at different noise uncertainty levels. The values of P_{MD} and P_{FA} are 10^{-5}, performance of SNR_{min} are then plotted in Fig. 2 with different noise uncertainty levels.

From the proposed method described in Sect. 3.1, it is clear that the proposed method is employed to improve SNR Wall by using transmitted signal S(t), which is distorted with channel noise n(t). Referring to the Fig. 1, the block diagram possesses two parts. The first part shows the cross-correlation method when signal is traveling through path1 and path2. At the receiver's end, there is cross-correlation which reduces interference noise. Considering the second part of the method, the energy detector improves SNR and detects unused

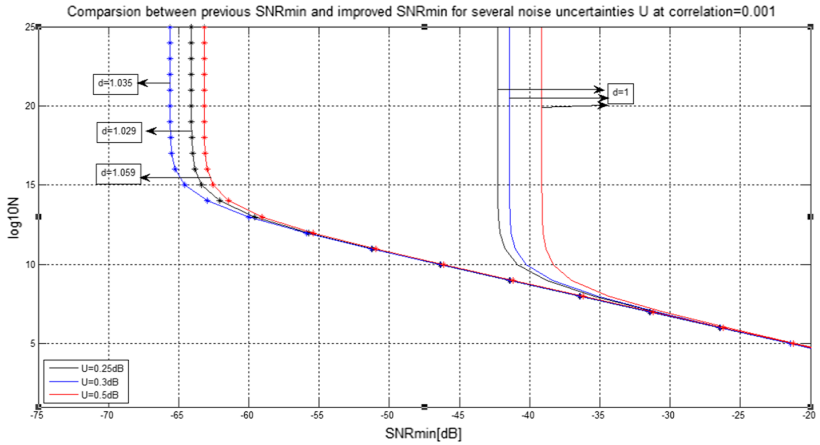


Fig. 2 Comparison of the Previous SNR_{min} [26, 28] and the Improved SNR_{min} for several noise uncertainties U and $\varepsilon = 0.001$

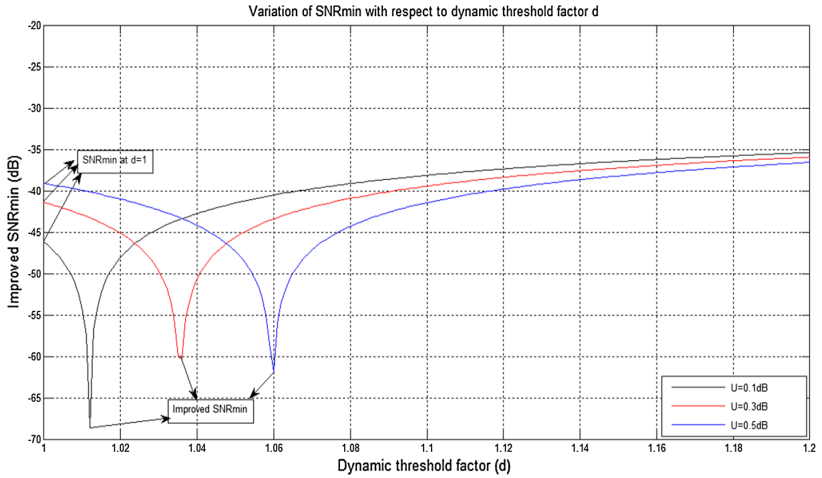


Fig. 3 Variation of SNR_{min} (dB) with respect to the dynamic threshold (d)

Table 1 Improved SNR_{min} and percentage of improvement by d at $U = 0.1$ dB

No.	SNR_{min} (d = 1)	Value of 'd'	Improved SNR_{min}	Improvement (%)
1.	− 46.327	1.008	− 51.478	11.11
2.	− 46.327	1.009	− 52.907	14.20
3.	− 46.327	1.010	− 55.044	18.81
4.	− 46.327	1.011	-59.404	28.22

primary user signals. Thus, at the output, SNR is the noise within band of interest. From Fig. 1, it can be seen that both are the real and imaginary parts. The reason is because, every samples are independent and identically distributed (IID), after that the central limit theorem for a large number of samples, their common converge is in the direction of a standard allocation. Equations (16–20) indicate the noise uncertainty, which is for a fixed threshold method. However, this degrades performance of detector. This is maintained by the constant in Eqs. (16, 18, 19) in such a way that the system performance is maintained constant to its previous desired value.

Figure 3 shows the SNR_{min} verses the number of samples at $U = 0.25$ dB, 0.3 dB, 0.5 dB and $\varepsilon = 0.001$. The comparison is expressed between SNR_{min} at $d = 1$ and the improved SNR_{min} for different values of d. We can observe that at $d = 1$, $SNR_{min} = -42$ dB is improved to -64 dB for $d = 1.029$. Figure 2 shows SNR_{min} with respect to the dynamic threshold factor 'd'. SNR_{min} can be obtained from Fig. 2. Clearly, SNR Wall is at -62 dB for $U = 0.5$ dB, and $f = 1.06$. The variation of SNR_{min} at various values of d and percentages of improvement with SNR_{min} at $d = 1$ is shown in Tables 1 and 2 ($U = 0.1$ dB and 0.5 dB respectively). These tables show percentage improvement in SNR_{min} for correlation noise factor $\varepsilon = 0.001$ and the number of samples $N = 10^{20}$.

Table 2 Improved $SNR_{\min}$ and percentage of improvement by d at $U=0.5$ dB

No.	$SNR_{\min}$ (d = 1)	Value of 'd'	Improved $SNR_{\min}$	Improvement (%)
1	− 39.135	1.03	− 42.395	8.33
2.	− 39.135	1.07	− 46.992	20.08
3.	− 39.135	1.08	− 44.196	12.93
4.	− 39.135	1.09	− 42.547	8.72

Table 1 shows the $SNR_{\min}$ verses the number of samples at $U=0.1$ dB and $\varepsilon = 0.001$. The comparison is expressed between $SNR_{\min}$ at $d=1$ and the improved $SNR_{\min}$ for different values of d. We can observe that at $d=1$, $SNR_{\min}=-46.327$ dB is improved to -59.404 dB for $d=1.011$ and the percentage of improvement is 28.22%. Table 2 shows the $SNR_{\min}$ verses the number of samples at $U=0.5$ dB and $\varepsilon = 0.001$. The comparison is expressed between $SNR_{\min}$ at $d=1$ and the improved $SNR_{\min}$ for different values of d. We can observe that at $d=1$, $SNR_{\min}=-39.135$ dB is improved to -46.992 dB for $d=1.07$ and the percentage of improvement is 20.08%.

Figure 4 shows the probability of misdetection $\left(P_{MD}=1-P_D\right)$ for $U=0$ dB and $SNR=10$ dB. We can observe that P_{MD} increases as noise correlation ε increases from 0 to 1. As $\varepsilon \rightarrow 1, P_{MD} \rightarrow 1$ or $P_D \rightarrow 0$ and system shows a zero performance. In this figure, P_{MD} is plotted for different values of ε at $U=0.5$ dB and $d=1.035$.

Figure 5 is drawn at $P_{FA,des}=0.1$ and $SNR=-10$ dB. We can see that lower ε is required for higher U for robust detection. Also, the introduced d factor improves performance of P_{MD}.

Figure 6 indicates the comparative analysis between the proposed method and the existing ones [3, 26, 28] in term of energy detection under the number of noise uncertainty. It is clearly indicated that the proposed work performs better by the rate of 20.08% when the values of N increases and $\varepsilon = 0.001$.

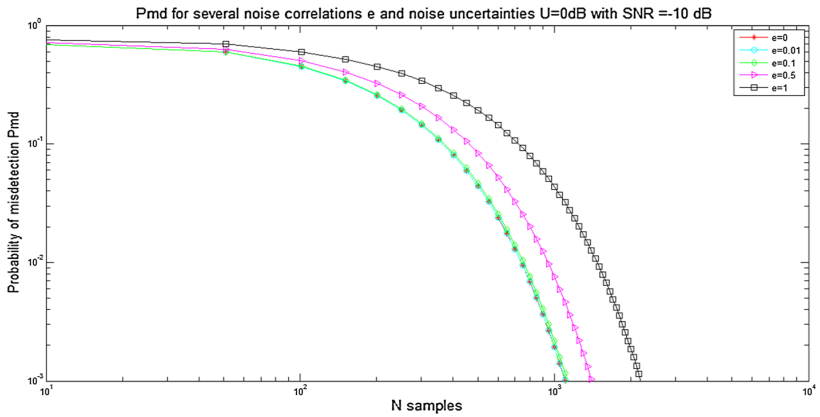


Fig. 4 Probability of misdetection $\left(P_{MD}\right)$ by samples (N) by values of ε

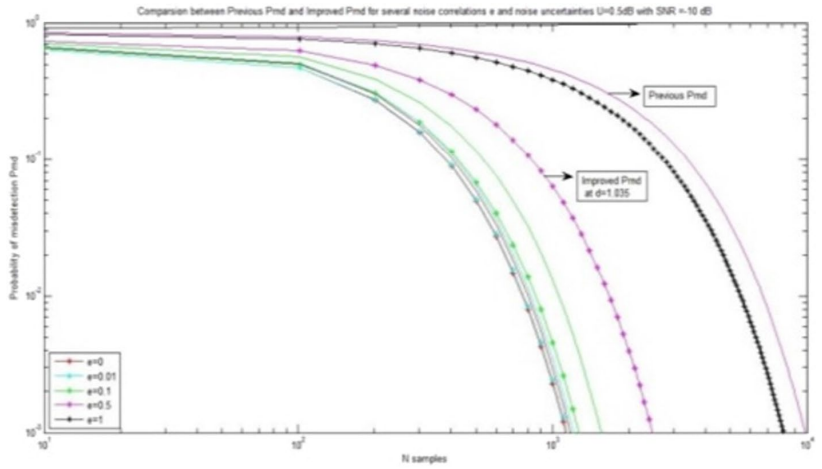


Fig. 5 Improved P_{MD} (at $d=1.035$) and the previous P_{MD} with $U=0.5$ dB, $SNR=-10$ dB [26, 28]. Improved P_{MD} at $d=1.035$. Previous P_{MD}

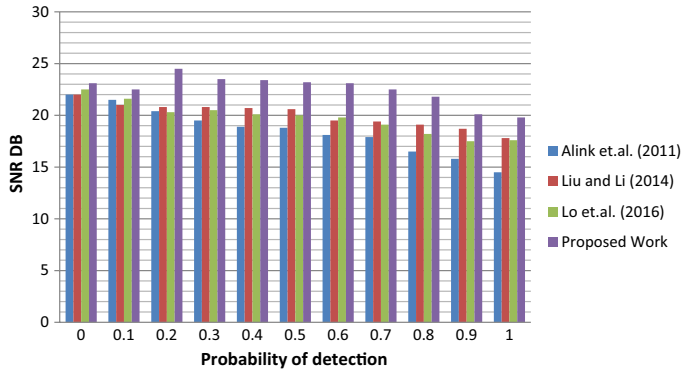


Fig. 6 Probability of energy detection under the number of noise uncertainty (N = 1000, ε = 0.001)

5 Conclusion

This article analyzed the performance of energy detector degraded when noise uncertainty is available. In this paper, we introduced a new hybrid method based on cross-correlation method and adaptive threshold method. An expression for SNR wall was derived which relates to correlation ε, uncertainty U and dynamic threshold factor d. In the cross-correlation, same signals pass to two different paths which reduce uncorrelated noises. Therefore, SNR is improved at outputs. Various simulations have been carried out to verify the accuracy of the derived relation on SNR degradation and packet loss charge. We identified

various noises (Eqs. 13 to 24). Uncertainty yields out to derive a lower bound on the SNR Wall of the sensor networks were investigated. The results obtained from the previous techniques were compared with the proposed method. Simulations showed that new derived SNR wall improves detection performance compares to the correlation-based method. The experiments clearly affirmed the advantages of the proposed work.

Future studies will investigate an effective computational model that integrates different data sources in the proposed method [8, 13, 16, 38, 39]; Kapoor et al. [21, 22, 29, 32–35, 40–54].

References

1. Ali, K. A., Rehmani, H., & Reisslein, M. (2016). Cognitive radio for smart grids: Survey of architectures, spectrum sensing mechanisms, and networking protocols. *IEEE Communications Surveys & Tutorials, 18*(1), 860–898.
2. Ali, G., Qaraqe, K., Celebi H. &Arslan H. (2010) An adaptive threshold method for spectrum sensing in multi-channel cognitive radio networks. In *2010 IEEE 17th international conference on telecommunications (ICT 2010)*, (pp. 425–429), IEEE.
3. Alink, M. S. O., Kokkeler, A. B., Klumperink, E. A., Smit, G. J., & Nauta, B. (2011). Lowering the SNR wall for energy detection using cross-correlation. *IEEE Transactions on Vehicular Technology, 60*(8), 3748–3757.
4. Beaulieu, D. L. (2000). Comprehensive reform and American Indian education. *Journal of American Indian Education, 39*(2), 29–38.
5. Bruckner, D., Velik, R., & Penya, Y. (2011). Machine perception in automation: A call toarms. *EURASIP Journal on Embedded Systems, 6*(8), 1–9.
6. Claudino, L., & Abrão, T. (2017). Spectrum sensing methods for cognitive radio networks: A review. *Wireless Personal Communications, 95*(4), 5003–5037.
7. Dhurgadevi, M., & Devi, P. M. (2018). An analysis of energy efficiency improvement through wireless energy transfer in wireless sensor network. *Wireless Personal Communications, 98*(4), 3377–3391.
8. Doss, S., Nayyar, A., Suseendran, G., Tanwar, S., Khanna, A., Son, L. H., et al. (2018). APD-JFAD: Accurate prevention and detection of Jelly Fish attack in MANET. *Ieee Access, 6,* 56954–56965.
9. Dutta P. & Manna G.C. (2016) Designing a cognitive radio with enhancement in throughput and improved spectrum sensing technique. In *2nd IEEE international conference on control science and systems engineering (ICCSS 2016)*, (pp. 27–29) July 2016, Singapore. https://doi.org/10.1109/ccsse.2016.7784345.
10. Emara, M., et al. (2016). Spectrum sensing optimization and performance enhancement of cognitive radio networks. *Wireless Personal Communications, 86*(2), 925–941.
11. Gao, N., et al. (2017). Robust collaborative spectrum sensing using PHY-layer fingerprints in mobile cognitive radio networks. *IEEE Communications Letters*. https://doi.org/10.1109/lcomm.2017.2656901.
12. Gao, R., Li, Z., Li, H., & Ai, B. (2015). Absolute value cumulating based spectrum sensing with Laplacian noise in cognitive radio networks. *Wireless Personal Communications, 83*(2), 1387–1404.
13. Garg, R, Mittal, M, Son, LH (2019). Reliability and energy efficient workflow scheduling in cloud environment. Cluster Computing, (**in press**).
14. Gohain, P. B., Chaudhari, S., &Koivunen, V. (2018) Cooperative energy detection with heterogeneous sensors under noise uncertainty: SNR wall and use of evidence theory. IEEE Transactions on Cognitive Communications and Networking.
15. Gunes, N., Higgins, D. M., & Leeson, S. M. (2016). A stochastic resonator to detect BPAM signals; analysis, PSR designs, and sine-induced SR. *IET Signal Processing*. https://doi.org/10.1049/iet-spr.2015.0152.
16. Hai, D. T., Son, H., & Vinh, L. T. (2017). Novel fuzzy clustering scheme for 3D wireless sensor networks. *Applied Soft Computing, 54,* 141–149.
17. Haijun, Z., et al. (2016). Interference-limited resource optimization in cognitive femtocells with fairness and imperfect spectrum sensing. *IEEE Transactions on Vehicular Technology, 65*(3), 1761–1771.
18. Jha, S. K., & Eyong, E. M. (2018). An energy optimization in wireless sensor networks by using genetic algorithm. *Telecommunication Systems, 67*(1), 113–121.

19. Joshi, D., Dimitrie, P., & Octavia, D. (2011). Gradient-based threshold adaptation for energy detector in cognitive radio systems. *IEEE Communications Letters, 15*(1), 19–21.
20. Junaid, I., & Kim, D. (2017). Energy-Efficient Management of Cognitive Radio Terminals with Quality-Based Activation. *IEEE Communications Letters, 21*(5), 1171–1174.
21. Kapoor, R., Gupta, R., Jha, S., & Kumar, R. (2018). Detection of power quality event using histogram of oriented gradients and support vector machine. *Measurement, 120,* 52–75.
22. Kapoor, R., Gupta, R., Kumar, R., Son, L. H., & Jha, S. (2019). New scheme for underwater acoustically wireless transmission using direct sequence code division multiple access in MIMO systems. *Wireless Networks*. https://doi.org/10.1007/s11276-018-1750-z.
23. Kapoor, R., Gupta, R., Son, L. H., Jha, S., & Kumar, R. (2018). Detection of power quality event using histogram of oriented gradients and support vector machine. *Measurement, 120,* 52–75.
24. Kapoor, R., Gupta, R., Son, L. H., Jha, S., & Kumar, R. (2018). Boosting performance of power quality event identification with KL Divergence measure and standard deviation. *Measurement, 126,* 134–142.
25. Kapoor, R., Gupta, R., Son, LH, Kumar, R., & Jha, S. (2018b) New scheme for underwater acoustically wireless transmission using direct sequence code division multiple access in MIMO systems. Wireless Networks, pp. 1–13.
26. Liu, J., & Li, Z. (2014). Lowering the signal-to-noise ratio wall for energy detection using parameter-induced stochastic resonator. *IET Communications, 9*(1), 101–107.
27. Liu, J., Youguo, W., & Qiqing, Z. (2016). Stochastic resonance of signal detection in mono-threshold system using additive and multiplicative noises. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 99*(1), 323–329.
28. Lo, Y. S., Lim, H. S., & Tan, A. W. C. (2016). Robust signal-to-noise ratio estimation in non-gaussian noise channel. *Wireless Personal Communications, 91*(2), 561–575.
29. Long, H. V., Ali, M., Khan, M., & Tu, D. N. (2019). A novel approach for fuzzy clustering based on neutrosophic association matrix. *Computers & Industrial Engineering*. https://doi.org/10.1016/j.cie.2018.11.007.
30. Mashreghi, M., & Abolhassani, B. (2017). A cluster-based cooperative spectrum sensing strategy to maximize achievable throughput. *Wireless Personal Communications, 96*(3), 4557–4584.
31. Oude, A., et al. (2011). Lowering the SNR wall for energy detection using cross-correlation. *IEEE Transactions on Vehicular Technology, 60*(8), 3748–3757.
32. Phuong, P. T. M., Thong, P. H., & Son, L. H. (2018). Theoretical analysis of picture fuzzy clustering: Convergence and property. *Journal of Computer Science and Cybernetics, 34*(1), 17–32.
33. Robinson, Y. H., Julie, E. G., Saravanan, K., Kumar, R., & Son, L. H. (2019). FD-AOMDV: fault-tolerant disjoint ad-hoc on-demand multipath distance vector routing algorithm in mobile ad-hoc networks. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-018-1126-3.
34. Saravanan, K., Anusuya, E., Kumar, R., & Son, L. H. (2018). Real-time water quality monitoring using Internet of Things in SCADA. *Environmental Monitoring and Assessment, 190*(9), 556.
35. Saravanan, K., Aswini, S., Kumar, R., & Son, L. H. (2019). How to prevent maritime border collision for fisheries?-A design of Real-Time Automatic Identification System. *Earth Science Informatics, 1,* 1–12. https://doi.org/10.1007/s12145-018-0371-5.
36. Sarkar, S., Virani, N., Yasar, M., Ray A. & Sarkar S. (2013) Spatiotemporal information fusion for fault detection in shipboard auxiliary systems. American Control Conference, Washington D. C., (pp. 3846–3851).
37. Shaikh B., Zafi S., &Umrani F. (2016) An unsigned autocorrelation based blind spectrum sensing approach for cognitive radio. In *2016 IEEE International Conference on Open Source Systems & Technologies (ICOSST 2016)*, 15–17 Dec. 2016, Lahore, Pakistan. https://doi.org/10.1109/icosst.2016.7838576.
38. Singh, K., Singh, K., Son, L. H., & Aziz, A. (2018). Congestion control in wireless sensor networks by hybrid multi-objective optimization algorithm. *Computer Networks, 138,* 90–107.
39. Singh, N., Son, L. H., Chiclana, F., & Jean-Pierre, M. (2019). A new fusion of salp swarm with sine cosine for optimization of non-linear functions. *Engineering with Computers*. https://doi.org/10.1007/s00366-018-00696-8.
40. Son, L. H. (2015). A novel kernel fuzzy clustering algorithm for geo-demographic analysis. Information Sciences—Informatics and Computer Science. *Intelligent Systems, Applications: An International Journal, 317,* 202–223.
41. Son, L. H. (2016). Generalized picture distance measure and applications to picture fuzzy clustering. *Applied Soft Computing, 46,* 284–295.
42. Son, L. H., & Hai, P. V. (2016). A novel multiple fuzzy clustering method based on internal clustering validation measures with gradient descent. *International Journal of Fuzzy Systems, 18*(5), 894–903.

43. Son, L. H., Jha, S., Kumar, R., Chatterjee, J. M., & Khari, M. (2019). Collaborative handshaking approaches between internet of computing and internet of things towards a smart world: a review from 2009 to 2017. *Telecommunication Systems*. https://doi.org/10.1007/s11235-018-0481-x.
44. Son, L. H., & Tien, N. D. (2017). Tune up fuzzy C-means for big data: some novel hybrid clustering algorithms based on initial selection and incremental clustering. *International Journal of Fuzzy Systems, 19*(5), 1585–1602.
45. Son, L. H., & Tuan, T. M. (2016). A cooperative semi-supervised fuzzy clustering framework for dental X-ray image segmentation. *Expert Systems with Applications, 46,* 380–393.
46. Son, L. H., & Thong, P. H. (2017). Some novel hybrid forecast methods based on picture fuzzy clustering for weather nowcasting from satellite image sequences. *Applied Intelligence, 46*(1), 1–15.
47. Son, L. H., & Tuan, T. M. (2017). Dental segmentation from X-ray images using semi-supervised fuzzy clustering with spatial constraints. *Engineering Applications of Artificial Intelligence, 59,* 186–195.
48. Tam, N. T., Hai, D. T., Son, L. H., & Vinh, L. T. (2018). Improving lifetime and network connections of 3D wireless sensor networks based on fuzzy clustering and particle swarm optimization. *Wireless Networks, 24*(5), 1477–1490.
49. Tandra, R., & Sahai, A. (2005). Fundamental limits on detection in low SNR under noise uncertainty. In *Wireless Networks, Communications and Mobile Computing, 2005 International Conference on* (Vol. 1, pp. 464–469). IEEE.
50. Thanh, N. D., Ali, M., & Son, L. H. (2017). A novel clustering algorithm in a neutrosophic recommender system for medical diagnosis. *Cognitive Computation, 9*(4), 526–544.
51. Thong, P. H., & Son, L. H. (2016). Picture fuzzy clustering: a new computational intelligence method. *Soft Computing, 20*(9), 3549–3562.
52. Thong, P. H., & Son, L. H. (2016). A novel automatic picture fuzzy clustering method based on particle swarm optimization and picture composite cardinality. *Knowledge-Based Systems, 109,* 48–60.
53. Thong, P. H., & Son, L. H. (2016). Picture fuzzy clustering for complex data. *Engineering Applications of Artificial Intelligence, 56,* 121–130.
54. Tuan, T. M., Ngan, T. T., & Son, L. H. (2016). A novel semi-supervised fuzzy clustering method based on interactive fuzzy satisficing for dental X-ray image segmentation. *Applied Intelligence, 45*(2), 402–428.
55. Zulfikar, A. et al. (2017) Enhanced spectrum sensing based on Energy detection in cognitive radio network using adaptive threshold. In *2017 IEEE international conference on networking, systems and security (NSysS)*, Doi: https://doi.org/10.1109/nsyss.2017.7885815.

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



Rajiv Kapoor received M.E. and Ph.D. degree in Electronics & Communication Engineering from Delhi College of Engineering, Delhi University. Dr. Kapoor presently working as Professor in Electronics & Communication Engineering Department, AIACT&R (Govt. of NCT of Delhi). He has authored over 90 research papers in various renowned international journal and conferences. His primary research interests are machine learning, computer vision, signal and image processing.



Rashmi Gupta received M.E. and Ph.D. degree in Electronics & Communication Engineering from Delhi College of Engineering, Delhi University. Dr. Gupta held the position of Sr. Engineer in Calcom group of companies from 1991 to 1999, position of Lecturer in Electronics & Communication Engineering Department at Hindu Institute of Technology and Senior Lecturer in Maharaja Agrasen Intitute of Technology, Delhi from 1999 to 2003. Dr. Gupta presently working as Associate Professor in Electronics & Communication Engineering Department, AIACT&R (Govt. of NCT of Delhi). She has authored over 40 research papers in various renowned international journal and conferences. Her primary research interests are machine learning, computer vision, signal and image processing.



Le Hoang Son obtained the PhD degree on Mathematics—Informatics at VNU University of Science, Vietnam National University (VNU). He has been promoted to Associate Professor in Information Technology since 2017. Currently, Dr. Son works as a Professor at Vietnam National University, Hanoi and a Visting Researcher at Ton Duc Thang University, Ho Chi Minh city, Vietnam. His major field includes Artificial Intelligence, Data Mining, Soft Computing, Fuzzy Computing, Fuzzy Recommender Systems, and Geographic Information System. He is a member of International Association of Computer Science and Information Technology (IACSIT), Center for Applied Research in e-Health (eCARE), Vietnam Society for Applications of Mathematics (Vietsam). Dr. Son serves as Editorial Board of International Journal of Ambient Computing and Intelligence (IJACI, SCOPUS), Editorial Board of Vietnam Journal of Computer Science and Cybernetics (JCC), Associate Editor of International Journal of Engineering and Technology (IJET), Associate Editor of Neutrosophic Sets and Systems (NSS), and Associate Editor of Vietnam Research and Development on Information and Communication Technology (RD-ICT).



Sudan Jha was born on October 22 at the Capital city Kathmandu of Himalayan Country Nepal. He received proficiency in certificate level from Saint Xavier's College Kathmandu and then Bachelor of Engineering in Electronics Engineering from Motilal Nehru Regional College, Allahabad. Uttar Pradesh, India, in 2001; joined as a Lecturer in one of the premium and largest Engineering College and the first one in the private domain in Nepal ("Nepal Engineering College (NEC)"), where he got full sponsorship from the employer (NEC) to pursue "Masters in Computer Science". In due course of time, he chaired and organized 5 International Conferences, some of the proceedings of those conferences been published by Springer Verlag, World Science Series, Imperial Press London, etc. He was promoted to Assistant Professor in the Department of Computer Science and Engineering after completion of his Masters Degree and no sooner than later, he became the Head of the Computer Science and Engineering Department.



Raghvendra Kumar is working as Assistant Professor in Computer Science and Engineering Department at L.N.C.T Group of College Jabalpur, M.P. India. He received B. Tech. in Computer Science and Engineering from SRM University Chennai (Tamil Nadu), India, M. Tech. in Computer Science and Engineering from KIIT University, Bhubaneswar, (Odisha) India and Ph.D. in Computer Science and Engineering from Jodhpur National University, Jodhpur (Rajasthan), India. He has published 86 research papers in international/National journal and conferences including IEEE, Springer and ACM as well as serve as session chair, Co-chair, Technical program Committee members in many international and national conferences and serve as guest editors in many special issues from reputed journals (Indexed By: Scopus, ESCI). He also received best paper award in IEEE Conference 2013 and Young Achiever Award-2016 by IEAE Association for his research work in the field of distributed database. His researches areas are Computer Networks, Data Mining, cloud computing and Secure Multiparty Computations, Theory of Computer Science and Design of Algorithms. He authored 12 computer science books in field of Data Mining, Robotics, Graph Theory, and Turing Machine by IGI Global Publication, USA, IOS Press Netherland, Lambert Publication, Scholar Press, Kataria Publication, Narosa, Edupedia Publication, S. Chand Publication and Laxmi Publication.

Annals of Data Science
https://doi.org/10.1007/s40745-018-00185-1

Applicability of Machine Learning in the Measurement of Emotional Intelligence

Manish Sharma[1] · Shikha N. Khera[1] · Pritam B. Sharma[2]

Received: 21 August 2018 / Revised: 23 November 2018 / Accepted: 13 December 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

The Trait Meta Mood Scale (TMMS) is one of the widely used instruments for measuring the emotional intelligence. This scale helps in ascertaining the overall emotional intelligence and can be used by organizations to handle the workforce and hence increase the efficiency and effectiveness by taking corrective measures, thereby transforming the organizations. If a large data set is available with some missing value, it becomes difficult to find the overall emotional intelligence of the given group and carry out the statistical analysis. This work proposes a model which applies neural network to find out the missing data and to perform regression. The model provides a flexible system to measure emotional intelligence. It paves a way for the application of machine learning in the TMMS scale of emotional intelligence but also in other scales of emotional intelligence.

Keywords Trait Meta Mood Scale · Neural networks · Regression · Machine learning · Emotional intelligence

1 Introduction

The Trait Meta Mood Scale (TMMS) is an instrument for measuring emotional intelligence (EI), based on the work by Salovey and Mayer [1]. It was developed in 1995 and was first of its kind. The developers extended the existing State Meta Mood Scale which was based on the ongoing moods. The scale contains 30 items rated on a 5 point scale. The response is based on the personal view point of the respondent. The TMMS measures the following factors of EI.

✉ Shikha N. Khera
drshikhakhera@gmail.com

1 Delhi Technological University, New Delhi, India

2 Amity University, Manesar, Haryana, India

Published online: 05 January 2019

(1) Attention to feeling and emotions
(2) Clarity in discrimination of feelings
(3) Mood repair

The first factor is measured by 13 items, second by 11 and the third by 6. A total of 15 items are reverse scored. Many researchers have pointed out the problems in TMMS, the most important of which is the scoring and is given by the person himself and hence is based on perception rather than the reality. Nevertheless, it is used for personal assessments and measuring psychosomatic disorders. One can't use this scale where interpersonal relations are more important than individual abilities, and the respondent has an incentive to impress.

In the process of data collection and subsequent analysis, it might happen that the data is incomplete, noisy or some of the fields have not been filled by the subject. In such cases the analysis becomes difficult. More so, in large organisations, this absence of data hinders the process of analysis and implementing subsequent corrective measures. In such cases, regression analysis can be used. There are many methods of carrying out this process. However, in the case of huge data, machine learning methods can be used and are effective and efficient. This work applies neural networks (NN), one of the machine learning (ML) methods, for carrying out the regression. The work can be extended for all such tests and therefore have wide applicability.

The work has been organised as follows. The introduction in the Sect. 1 is followed by the literature review in the Sect 2. Section 3 gives a brief overview of regression using neural networks; Sect. 4 explains the experiment as well as the results and the last section concludes.

2 Literature Review

The literature review was carried out in accordance with the guidelines proposed by Kitchenham [2] to find the techniques of measuring emotional intelligence. Since the present work aims to apply ML to EI, it is essential to find the most common tools for measuring EI. The following table (Table 1) presents some of the most important tools.

In the review, it was found that the techniques have some limitations. For instance, some of the scales are not designed as per the sociological conditions of the third world countries. Some are based on the self-appraisal, which may not help to attain the accurate results. As per the review, some of the scales cannot be used where interpersonal relations are important. However, the work can be done only on the existing scales. Khasawneh [3] investigated emotional intelligence influence on technology acceptance using the TMMS scale. Palomba [4] used TMMS to measure perceived EI of high brand loyalty video game players. Gutiérrez-Cobo et al. [5] used the Spanish translation of the TMMS as a self-report ability test, and Brito-Costa et al. [6] analyzed the psychometric properties of the Portuguese version of TMMS. The applications of Emotional Intelligence in the business are covered by the set of books by Harvard Business review [7].

Table 1 Literature review of the emotional intelligence tools

Name of the Scale	References	Structure
Mayer-Salovey-Caruso Emotional Intelligence Test [8]	John Mayer, Peter Salovey and David Caruso (2002)	The scale has 141 items. The items are grouped into various tasks. These are further divided into categories
Trait Meta-Mood Scale [1]	Salovey, P., Mayer, J. D., Goldman, S. L., Turvey, C., & Palfai, T. P. (1995)	The scale has 30 items. The answers are on a five point Likert scale ranging from 1 to 5
Wong and Law Emotional Intelligence Scale [9]	Chi-Sum Wong and Kenneth S. Law in 2002	The scale has 16 items to be rated on a 7-point Likert scale ranging from 1 to 7
Wong's Emotional Intelligence Scale [10]	Wong et al. (2004)	The WEIS consists of 40 items divided into two parts of 20 items each
Tapia Emotional Intelligence Inventory [11]	Martha Tapia (2001)	It consists of 41 items. There are four subscales
Schutte self report emotional intelligence test [12]	Nicola Schutte (1998)	It consists of 33 items rated on a 5-point likert scale ranging from 1 (strongly dislike) to 5 (strongly like)
Emotional Quotient Inventory [13]	Reuven Bar-On (1997)	This test consists of 133 items
Group emotional intelligence survey [14]	Vanessa UrchDruskat and Steven B. Wolff (2003)	The test contains 57 items that measure the nine dimensions
Myers-Briggs Type Indicator [15]	Isabel Briggs Myers and Katharine Cook Briggs(1942)	The test consists of 93 questions. The choices are paired.
Belbin Team Role Inventory [16]	Meredith Belbin (1993)	The test has a lot of staconsist of series of statement each followed by nutements followed by responses. The subject assigns a point to the references
Leadership Practice Inventory [17]	James M. Kouzes and Barry Z Posner(1987)	There are three varieties of LPI namely Lpi 360, lPI individual and LPI leader
Multifactor Leadership Questionnaire [18]	Bernard M. Bass and Bruce J. Avolio (1995)	The scale consists of 45 items based on the 5-point frequency scale
Rotterdam Emotional Intelligence Scale (REIS) [19]	Pekaar, K. A., Bakker, A. B., van der Linden, D. and Born, M. P., (2018)	It consists of 28 items based on a 5-point Likert scale

Table 2 Terminology of machine learning

Inductive bias	The set assumptions we make to have learning possible is called inductive bias [21]
Generalization	How well a model trained on the training set predicts the right output for new instances is called generalization [21]
Under fitting:	For best generalization, we should match the complexity of the hypothesis class H with the complexity of the function underlying the data. If H is less complex than the function, we have underfitting [21]
Over fitting	In the above case, if H is more complex than the function, we have underfitting
Validation set	From the given training data, one part is used for training, and the other part is used to test the generalization ability of the model [21]
Cross validation	From among the given hypothesis, the most accurate on the validation set is the best one. The process is referred to as cross validation [21]
Test set	The model is tested with the unseen data, and its accuracy, false positive and false negative are measured in a phase called testing. This unseen data is test set

3 Machine Learning and Regression Using Neural Network

ML techniques aim at learning from a training set and applying the model to a yet unseen test set. The learning can be supervised or unsupervised. In supervised learning, labels are provided along with the test set whereas in case of unsupervised learning no labels are provided. In the case of supervised learning, the teacher tries to minimize the error between what is obtained from the model and the target. Therefore, two major things are required, the learning algorithm and the measure by which error will be calculated. ML is used in a wide variety of problems like image processing, medical diagnosis etc. Here, learning means using the past experience to enhance future predictions. Two major problems dealt with ML are classification and regression. The task of a classifier component is to use the feature vector provided by the feature extractor to assign the object to a category [20]. In regression, we seek to find some functional description of the data often with the goal of predicting the values of a new input [20]. Unsupervised learning, on the other hand, does not require a teacher and solves problems by clustering. In Clustering there is no explicit teacher, and the system forms clusters or "natural groupings" of input patterns [20]. Another type of learning is reinforcement learning wherein the goodness of the policies and past actions reaffirm what is being produced. The reinforcement learning refers to the most typical way to train a classifier is to present an input, compute its tentative category label, and use the known target category label to improve the classifier [20].

In supervised learning, if the output is discrete then labelled classes can be obtained. Whereas, in the case of continuous output the machine becomes capable of carrying the regression analysis. This work performs regression analysis. Table 2 briefly introduces the terminology pertinent to ML and frequently used in the paper.

In machine learning, the training data consist of features and labels. The value of labels depends on features. In this context, the feature can be considered as independent variables and labels are dependent variables. It becomes important to find the strength of the relationship between dependent variables and independent variables which is

referred to as regression [22]. Statistical regression is a set of procedures for estimating the relationship among variable [23]. It tells us how the value of one variable changes with the other. The technique attempts to find the expression of the dependent variable in term of the independent variable, which is the regression function. The technique has been used for predicting unknown values. From among many statistical techniques, the Least Square Method (LSM) for finding the unknown coefficient is one of the most common methods. Lately, ML techniques like NN are being used to carry out the analysis.

In the NN model, if the training data is (X, Y) where,
$X = \{X_1, X_2, X_3, \ldots\ldots, X_n\}$
$Y = \{Y_1, Y_2, Y_3 \ldots\ldots, Y_n\}$, where X is the dependent variable(s) and Y is the independent variable
The technique attempts to find W, such that
$Y = WX + W_0$, where Y and X are matrices.

In the NN model, an additional input $x1 = 1$ can be taken to account for the bias. In linear perception, the output is W.X., which needs to be corrected by Gradient Decent. The Gradient Decent method seeks to minimize the mean squared error. The minimum value of the function can be reached by differentiating E w.r.t W and subsequently putting the derivative equal to 0. In the formula that finds the weight in the next iteration, the learning rate is incorporated. If the value of the coefficient representing the learning rate is too small, the conversion would be slow. On the other hand, if the value is too large, there can be a lot of oscillations. This rule is generally referred to as Delta rule, Gradient rule or the Root Mean Square Technique. This technique guarantees the best possible solution but is remarkably slow.

4 Experiment and Results

The purpose of the experiment was to find out the unknown data in a given sample of TMMS data. The process was carried by training a neural network by data collected from 38 subjects. The data had 22 fields, and the target data is the total TMMS data scored. It may be noted that the TMMS score is calculated by giving +1 weight to some of the fields and reverse to the others. 70% of the data was used for training model and 15% for validation. The validation data helps in making a better model in the process of training. The test data is used to ascertain the model. The test data is not seen by the model and hence the results obtained by the test data establish the correctness of the proposed model. In our model, 15% of the data was used for testing. The number of hidden neurons in the model was 10. This has been depicted in Fig. 1.

The data was collected by choosing 38 persons aged between 25 and 35. The sample chosen was representative of the population as the ethos of the sample was the same as that of the desired population. Moreover, the mean age of the population was 31, and the standard deviation of the ages is 4.6. Utmost care was taken in gathering the data and analyzing it. The mean of the final score was 43.34 and the standard deviation of 8.811. The data and the list of participants have been preserved for replicating the experiments

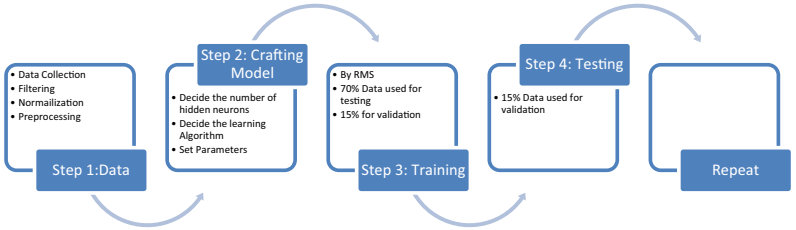


Fig. 1 Regression using neural networks

and for the extension of the work. In our study, 38 subjects were considered. In our work, we use both total and partial score to predict the classification.

The data was pre-processed and then subjected to the experiment. First of all missing values were filled as follows. If for a particular person, the majority of the fields were incomplete, his data was not taken as the training example of the model. In other cases, regression analysis was applied to fill the unknown values. The data was then normalized by calculating the z score

$$z = \frac{x - m}{s},$$

where x is the value, m is the mean and s is the standard deviation. However, later it was realized that the normalization did not have any effect on the results. The data collected had 22 features.

In the proposed model, the mean squared error was taken as the one to be minimised. The mean squared error is the distance between output and target [21]. The process was carried out 100 times, and the average mean squared error was 0.746562. The standard deviation obtained was 1.799185 which established the ability of the model to find out the predicted values correctly. The experiment was carried out 50 times, and the average was 0.73, and the standard deviation was 7.6. The process is depicted in the figure below (Fig. 1). Table 3 shows the results obtained in one of the trails and the corresponding variation in the root mean square error has been depicted in Table 4 and Fig. 2.

5 Conclusion and Future Scope

Regression analysis is used to find the relation between the dependent and the independent variables. In the data obtained by the EI tools, this process can be used to find the missing data and hence contribute in better and effective assessment. The process can be carried out using statistical methods, but in the case of large data, the methods would not be efficient. This work applies ML methods to accomplish the said task. NN has been used to carry out the regression. The model has been implemented and has been tested with the data collected. The experiment has been repeated one hundred times, and the average mean squared error was 0.746562. The standard deviation obtained was 1.799185. The results suggest that the technique is not only effective but

Table 3 The results obtained in one of the trails

Target	TMMS score	Mean square error
33	33.07518	0.005652
58	58.59404	0.352881
43	39.77965	10.37068
52	52.02129	0.000453
39	39.57002	0.324927
48	48.40801	0.166469
53	53.23572	0.055565
48	48.72537	0.526163
36	36.26308	0.069211
39	39.88064	0.775524
37	38.13127	1.279762
54	54.52121	0.27166
49	49.73531	0.540685
47	46.78122	0.047865
35	35.89181	0.795325
56	56.75487	0.569836
55	55.4225	0.178503
41	41.69184	0.478644
36	36.97349	0.947674
29	28.89121	0.011835
48	48.61453	0.377652
46	46.73448	0.539462
41	41.27804	0.077304
42	42.30728	0.094422
47	47.78121	0.610289
27	25.00433	3.982692
31	31.2973	0.088387
46	46.26559	0.070536
53	52.51211	0.238035
51	51.28975	0.083954
23	23.64396	0.41468
43	41.91359	1.180277
52	51.76181	0.056734
48	48.75302	0.56704
40	40.48606	0.236258
44	43.8889	0.012343
49	48.24351	0.572272
28	29.78646	3.191434

Table 4 The summary of the results

Sum	30.16308
Average	0.793765
SD	1.786534

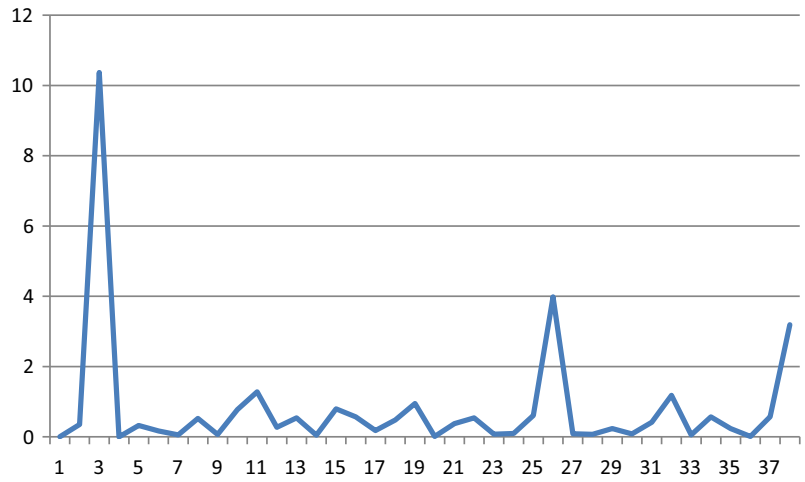


Fig. 2 Variation in the root mean square error

is also efficient as compared to the conventional methods. In the future, machine learning may be used for tests used to measure emotional intelligence other than TMMS. This paper is useful to the organizations who measure the emotional intelligence of a number of people.

References

1. Salovey P, Mayer JD, Goldman SL, Turvey C, Palfai TP (1995) Emotional attention, clarity, and repair: Exploring emotional intelligence using the Trait Meta-Mood Scale. American Psychological Association, Washington
2. Kithenham BA (2012) Systematic review in software engineering: where we are and where we should be going. In: Proceedings of the 2nd international workshop on evidential assessment of software technologies, pp 1–2
3. Khasawneh OY (2018) Technophobia without boarders: the influence of technophobia and emotional intelligence on technology acceptance and the moderating influence of organizational climate. Comput Hum Behav 88:210–218
4. Palomba A (2018) Virtual perceived emotional intelligence: how high brand loyalty video game players evaluate their own video game play experiences to repair or regulate emotions. Comput Hum Behav 85:34–42
5. Gutiérrez-Cobo MJ, Cabello R, Fernández-Berrocal P (2017) Performance-based ability emotional intelligence benefits working memory capacity during performance on hot tasks. Sci Rep 7(1):11700
6. Brito-Costa S, Castro FV, Moisao A, Alberty A, de Almeida H, Fernández MIR (2016) Psychometric properties of Portuguese version of trait meta-mood scale (TMMS24). Int J Dev Educ Psychol (Revista INFAD de Psicología) 2(1):133–142

7. Goleman D, McKee A, George B, Ibarra H (2018) HBR Emotional intelligence boxed set (6 Books) (HBR emotional intelligence series). Harvard Business Press, Brighton
8. Mayer JD, Salovey P, Caruso DR, Sitarenios G (2003) Measuring emotional intelligence with the MSCEIT V2.0. Emotion 3:97–105
9. Wong CS, Wong PM, Law KS (2007) Evidence on the practical utility of wong's emotional intelligence scale in Hong Kong and mainland China. Asia Pac J Manag 24:43–60
10. Law KS, Wong CS, Song LJ (2004) The construct and criterion validity of emotional intelligence and its potential utility for management studies. J Appl Psychol 89(3):483–496
11. Tapia M (2001) Measuring emotional intelligence. Psychol Rep 88(2):353-64s
12. Schutte NS, Malouff JM, Hall LE, Haggerty DJ, Cooper JT, Golden CJ et al (1998) Development and validation of a measure of emotional intelligence. Personal Individ Differ 25:167–177
13. Bar-On R (2004) The bar-on emotional quotient inventory (EQ-i): Rationale, description and psychometric properties. In: Geher G (ed) Measuring emotional intelligence: common ground and controversy. Nova Science, Hauppauge
14. Hamme C (2003) Group emotional intelligence: the research and development of an assessment instrument. Dissertation, Rutgers, New Brunswick, NJ
15. Myers IB, Myers PB (1995)[1980] Gifts differing: understanding personality type Mountain view. Davies-Black Publishing, California
16. Belbin R (1993) A reply to the belbin team-role self-perception inventory by furnham, Steele and Pendleton. J Occup Organ Psychol 66(3):259–260
17. Posner BZ, Kouzes JM (1988) Development and validation of the leadership practices inventory. Educ Psychol Measur 48(2):483–496
18. Bass BM, Aviolo BJ (1990) Multifactor leadership questionnaire. Consulting Psychologists Press, Palo Alto
19. Pekaar KA, Bakker AB, van der Linden D, Born MP (2018) Self-and other-focused emotional intelligence: development and validation of the rotterdam emotional intelligence scale (REIS). Person Individ Differ 120:222–233
20. Duda OR (2015) Pattern classification. Wiley, Hoboken
21. Alpaydin E (2017) Introduction to machine learning. MIT Press, Cambridge
22. Zureda JM (1996) Introduction to artificial neural system. West Publishing Company, Eagan
23. Wikipedia. Regression analysis. https://en.wikipedia.org/wiki/Regressionanalysis. Accessed 15 Apr 2018

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

RACSAM
https://doi.org/10.1007/s13398-019-00633-4

ORIGINAL PAPER

Approximation by genuine Gupta–Srivastava operators

Ram Pratap[1] · **Naokant Deo**[1]

Received: 25 June 2018 / Accepted: 14 January 2019
© The Royal Academy of Sciences, Madrid 2019

Abstract
In the present paper, we consider new operators, which is defined by Gupta and Srivastava (Eur J Pure Appl Math 11(3):575–579, 2018). They considered a general sequence of positive linear operators and gave the modified form of their previous operators (Neer et al. in Math Comput Model 37:1307–1315, 2003). As these operators preserve linear functions, we call these operators as genuine Gupta–Srivastava operators. Here we discuss some basic properties, direct results and rate of convergence of functions of bounded variation and weighted approximation.

Keywords Baskakov operators · Hypergeometric function · Function of Bounded variation

Mathematics Subject Classifications 41A25 · 42A36

1 Introduction

In [16], Srivastava and Gupta proposed a general sequence of positive linear operators and studied the rate of convergence for the functions of bounded variation. Two years later Ispir and Yüksel [13] considered the Bézier variant of the operators [16] and established the rate of convergence of the Bézier variant of Srivastava–Gupta operators. The operators [16] reproduced constant function only except for the special case $c = 0$, which provides the well-known Phillips operators. These operators and related versions were studied by several researchers, see for instance [3,4,9–11,13,15,18,19], where many interesting results, like the rate of convergence for functions having derivatives of bounded variation, Voronovskaja type asymptotic formula, better estimates and simultaneous approximation of these operators were established. It is well-known that if operators preserve the linear function, one may get a better approximation. In this direction very recently Gupta and Srivastava [17] proposed a general family of positive linear operators, which preserve constant as well as linear functions

✉ Naokant Deo
naokantdeo@dce.ac.in

Ram Pratap
rampratapiitr@gmail.com

[1] Department of Applied Mathematics, Formerly Delhi College of Engineering, Delhi Technological University, Bawana Road, Delhi 110042, India

Published online: 28 January 2019

for all $c \in \mathbb{N} \cup \{0\} \cup \{-1\}$, which may be termed as Gupta–Srivastava operators and for all integers m are defined as:

$$L_{n,c}(f(t);x) = \{n + (m+1)c\} \sum_{k=1}^{\infty} p_{n+mc,k}(x;c) \int_0^{\infty} p_{n+(m+2)c,k-1}(t;c) f(t)dt + p_{n+mc,0}(x;c) f(0), \tag{1.1}$$

where

$$p_{n,k}(x;c) = \frac{(-x)^k}{k!} \phi_{n,c}^{(k)}(x).$$

The special cases of operators (1.1) are as follows:

(i) For $c = 0$ and $\phi_{n,0}(x) = e^{-nx}$, we get Phillips operators

$$L_{n,0}(f(t);x) = n \sum_{k=1}^{\infty} p_{n,k}(x;0) \int_0^{\infty} p_{n,k-1}(t;0) f(t)dt + p_{n,0}(x;0) f(0),$$

where

$$p_{n,k}(x,0) = \frac{e^{-nx}(nx)^k}{k!} \quad \text{and} \quad x \in [0,\infty).$$

(ii) For $c \in \mathbb{N}$ and $\phi_{n,c}(x) = (1+cx)^{-\frac{n}{c}}$, we get genuine Baskakov–Durrmeyer type operators. These operators are similar to (1.1), where

$$p_{n,k}(x;c) = \frac{\left(\frac{n}{c}\right)_k}{k!} \frac{(cx)_k}{(1+cx)^{\frac{n}{c}+k}} \quad \text{and} \quad x \in [0,\infty),$$

with $(n)_i$ denotes the rising factorial given by

$$(n)_i = n(n+1)(n+2)...(n+i-1) \ \ \& \ \ (n)_0 = 1 (i \in \mathbb{N}).$$

(iii) For $c = -1$ and $\phi_{n,-1}(x) = (1-x)^{-n}$, we have sequence of Bernstein–Durrmeyer operators

$$L_{n,-1}(f,x) = (n-m-1) \sum_{k=1}^{n-m-1} p_{n-m,k}(x,-1) \int_0^1 p_{n-m-2,k-1}(t,-1) f(t)dt + p_{n-m,0}(x,-1) f(0) + p_{n-m,n-m}(x,-1),$$

where

$$p_{n,k}(x;-1) = \binom{n}{k} x^k (1-x)^{n-k} \quad \text{and} \quad x \in [0,1].$$

Recently Acu and collaborators [1,2] and Deo et al. [5] studied several interesting approximation properties and provided a better degree of approximation over the classical operators for instance direct approximation theorem using the Ditzian–Totik modulus of smoothness and a Voronovskaja type theorem.

In the present paper, we consider operators (1.1) and study fundamental properties, the rate of convergence in terms of second-order modulus of continuity, Lipschitz type space, Voronovskaya type estimates, convergence estimates for the function having derivatives of bounded variation and weighted approximation.

2 Basic results

In this section, we prove some basic results which are useful to prove several theorems and results.

Lemma 2.1 *Let $f(t) = t^i$, $i = 0, 1, 2, 3, 4$ and $c \in \mathbb{N} \cup \{0\} \cup \{-1\}$, then we have*

(i) $L_{n,c}(1; x) = 1$;
(ii) $L_{n,c}(t; x) = x$;
(iii) $L_{n,c}\left(t^2; x\right) = \frac{(n+(m+1)c)}{(n+(m-1)c)}x^2 + \frac{2}{(n+(m-1)c)}x$;
(iv) $L_{n,c}\left(t^3; x\right) = \frac{(n+(m+1)c)(n+(m+2)c)}{(n+(m-1)c)(n+(m-2)c)}x^3 + \frac{6(n+(m+1)c)}{(n+(m-1)c)(n+(m-2)c)}x^2 + \frac{6}{(n+(m-1)c)(n+(m-2)c)}x$;
(v) $L_{n,c}\left(t^4; x\right) = \frac{(n+(m+1)c)(n+(m+2)c)(n+(m+3)c)}{(n+(m-1)c)(n+(m-2)c)(n+(m-3)c)}x^4 + \frac{12(n+(m+1)c)(n+(m+2)c)}{(n+(m-1)c)(n+(m-2)c)(n+(m-3)c)}x^3 + \frac{36(n+(m+1)c)}{(n+(m-1)c)(n+(m-2)c)(n+(m-3)c)}x^2 + \frac{24}{(n+(m-1)c)(n+(m-2)c)(n+(m-3)c)}x$.

Proof All the moments of operators (1.1), can be obtained in terms of hypergeometric function of order $r \in \mathbb{N}$ for details see [17]. □

Remark 2.1 The central moment of the operators (1.1) are given as:

$$\mu_{n,c,s}(x) = L_{n,c}((t-x)^s; x),$$

for $s = 1, 2, 4$, we have

(i) $\mu_{n,c,1}(x) = 0$;
(ii) $\mu_{n,c,2}(x) = \frac{2x(1+cx)}{(n+(m-1)c)}$;
(iii) $\mu_{n,c,4}(x) = \frac{12c^2(n+(m+7)c)}{(n+(m-1)c)(n+(m-2)c)(n+(m-3)c)}x^4 + \frac{24c^2(13n+(13m+1)c)}{(n+(m-1)c)(n+(m-2)c)(n+(m-3)c)}x^3 + \frac{12(n+(m+9)c)}{(n+(m-1)c)(n+(m-2)c)(n+(m-3)c)}x^2 + \frac{24}{(n+(m-1)c)(n+(m-2)c)(n+(m-3)c)}x$.

Remark 2.2 For $n \in \mathbb{N}$, we have

$$L_{n,c}\left((t-x)^2; x\right) \leq \frac{2\psi_n^2(x)}{n},$$

where $\psi_n^2(x) = x(1+cx)$.

Lemma 2.2 *For $f \in C_B[0, \infty)$ and sufficient large n, we have*

(i) $\lim_{n\to\infty} n\mu_{n,c,2}(x) = 2x(1+cx)$;
(ii) $\lim_{n\to\infty} n^2\mu_{n,c,4}(x) = 12x^2(c^2x^2 + 26c^2x + 1)$.

Lemma 2.3 *For $f \in C_B[0, \infty)$, we have*

$$\left|L_{n,c}(f; x)\right| \leq \|f\|.$$

Theorem 2.1 *Let $f \in C_B[0, \infty)$ and $n \to \infty$. Then $\lim_{n\to\infty} L_{n,c}(f(t); x) = f(x)$, uniformly in each compact subset of $[0, \infty)$.*

Proof From Lemma 2.1, $L_{n,c}(1; x) = 1$, $L_{n,c}(t; x) = x$, $L_{n,c}(t^2; x) = x^2$, as $n \to \infty$. By Bohman–Korovkin theorem, we have

$$L_{n,c}(f(t); x) = f(x), \quad \text{as } n \to \infty,$$

uniformly in each compact subset of $[0, \infty)$. □

3 Main results

Now we shall give the main results in this section.

Let $C_B[0, \infty)$ be the space of all the real valued continuous and bounded functions f on the interval $[0, \infty)$, equipped with the norm

$$\|f\| = \sup_{x\in I} |f(x)|.$$

For $f \in C_B[0, \infty)$, The Peetre's K-functional is defined as:

$$K_2(f; \delta) := \inf_{x\in W_\infty^2} \left\{\|f - g\| + \delta \left\|g''\right\|\right\}, \ \delta > 0,$$

where $W_\infty^2[0, \infty) = \left\{g \in C_B[0, \infty) : g', g'' \in C_B[0, \infty)\right\}$. By DeVore and Lorentz ([6], p.177, Theorem 2.4) there exists an absolute constant $C > 0$ such that

$$K_2(f; \delta) \leqslant C\omega_2\left(f; \sqrt{\delta}\right), \tag{3.1}$$

where $\omega_2\left(f; \sqrt{\delta}\right)$ is second order modulus of continuity defined by

$$\omega_2\left(f; \sqrt{\delta}\right) = \sup_{0<h<\sqrt{\delta}} \sup_{x\in[0,\infty)} |f(x+2h) - 2f(x+h) + f(x)|.$$

The usual modulus of smoothness (or simply modulus of continuity) for $f \in C_B[0, \infty)$ is given by

$$\omega(f; \delta) = \sup_{0<h\leqslant\delta} \sup_{x\in[0,\infty)} |f(x+h) - f(x)|.$$

Theorem 3.1 *For the function $f \in C_B[0, \infty)$, we have*

$$\left|L_{n,c}(f(t); x) - f(x)\right| \leq C\omega_2(f; \sqrt{\delta_n(x)}),$$

where $\delta_n(x) = \frac{2x(1+cx)}{(n+(m-1)c)}$.

Proof Let the function $g \in W_\infty^2[0, \infty)$. Using Taylor's expression, we have

$$g(t) = g(x) + g'(x)(t-x) + \int_x^t (t-u)g''(u)du. \tag{3.2}$$

Applying $L_{n,c}(.; x)$ on both side of the expression (3.2), we have

$$L_{n,c}(g(t); x) - g(x) = g'(x)L_{n,c}((t-x); x) + L_{n,c}\left(\int_x^t (t-u)g''(u)du; x\right).$$

Using Remark 2.1, we have

$$L_{n,c}(g(t); x) - g(x) = L_{n,c}\left(\int_x^t (t-u)g''(u)du; x\right) \tag{3.3}$$

and we know that

$$\int_x^t (t-u)g''(u)du \leq (t-x)^2 \left\|g''\right\|. \tag{3.4}$$

From (3.3) and (3.4) and Remark 2.1, we have

$$\left|L_{n,c}\left(g(t);x\right)-g(x)\right| \leq L_{n,c}((t-x)^2;x)\left\|g''\right\| = \frac{2x(1+cx)}{(n+(m-1)c)}\left\|g''\right\|. \tag{3.5}$$

From Lemma 2.3, we have

$$\left|L_{n,c}\left(f(t);x\right)\right| \leqslant \|f\|. \tag{3.6}$$

Hence from (3.5) and (3.6), we get

$$\begin{aligned}\left|L_{n,c}(f(t);x)-f(x)\right| &\leq \left|L_{n,c}(f(t)-g(t);x)-(f(x)-g(x))\right|+\left|L_{n,c}(g(t);x)-g(x)\right| \\ &\leq 2\|f-g\|+\frac{2x(1+cx)}{(n+(m-1)c)}\left\|g''\right\|.\end{aligned}$$

Taking infimum on the right hand side of the above inequlity for all $g \in W_{\infty}^2[0,\infty)$ and using (3.1), we get the desired result. □

We estimate the rate of convergence of the operators $L_{n,c}$ by means of the Lipschitz class $LipM(\alpha)$, $0<\alpha\leq 1$. We say that $f \in C_B[0,\infty)$ belongs to $LipM(\alpha)$ if the inequality

$$|f(t)-f(x)| \leq M|t-x|^{\alpha}, \tag{3.7}$$

holds.

Theorem 3.2 *If $f \in LipM(\alpha)$, $x\in[0,\infty)$ and $n>(m-1)c$, we have*

$$\left|L_{n,c}(f(t);x)-f(x)\right| \leq M(\delta_n(x))^{\frac{\alpha}{2}}.$$

Proof Since $f \in LipM(\alpha)$, $x\in[0,\infty)$ and from inequality (3.7) and applying the Holder's inequality with $p=2/r$ and $q=2/(2-r)$, we have

$$\begin{aligned}\left|L_{n,c}(f(t);x)-f(x)\right| &\leq L_{n,c}(|f(t)-f(x)|;x) \\ &\leq ML_{n,c}(|t-x|^{\alpha};x) \\ &\leq M(L_{n,c}((t-x)^2;x)^{\alpha/2} \\ &\leq M(\delta_n(x))^{\alpha/2},\end{aligned}$$

which is the required result. □

Below we present the Voronovskaya type Theorem for Gupta–Srivastava operators:

Theorem 3.3 *Let $f \in C_B[0,\infty)$ and if there exists second derivatives of function f at a fixed point $x\in[0,\infty)$, then we have*

$$\lim_{n\to\infty} n\left[L_{n,c}(f;x)-f(x)\right] = x(1+cx)f''(x).$$

Proof By the Taylor's series expansion, we have

$$f(t) = f(x)+(t-x)f'(x)+\frac{(t-x)^2}{2!}f''(x)+\varepsilon(t,x)(t-x)^2, \tag{3.8}$$

where $\varepsilon(t,x)\to 0$ as $t\to x$.

Applying $L_{n,c}(;x)$ on both side of (3.8), we have

$$\begin{aligned}L_{n,c}(f(t)-f(x);x) =& f'(x)L_{n,c}((t-x);x)+\frac{f''(x)}{2!}L_{n,c}((t-x)^2;x) \\ & L_{n,c}(\varepsilon(t,x)(t-x)^2;x).\end{aligned}$$

From Remark 2.1 $L_{n,c}((t-x);x)=0$,

$$\begin{aligned}\lim_{n\to\infty} nL_{n,c}(f(t)-f(x);x) =& \lim_{n\to\infty} nL_{n,c}((t-x)^2;x)f''(x) \\ &+\lim_{n\to\infty} nL_{n,c}(\varepsilon(t,x)(t-x)^2;x).\end{aligned} \tag{3.9}$$

From Theorem 2.1, Lemma 2.2 and applying Cauchy–Schwarz inequality in the last term of (3.9), we get

$$\lim_{n\to\infty} nL_{n,c}(\varepsilon(t,x)(t-x)^2;x)=0. \tag{3.10}$$

Using Lemma 2.2 and (3.9), (3.10), we get the required result. □

Let $B_2[0,\infty)$ be the space of all functions defined on $[0,\infty)$ given by

$$B_2[0,\infty) = \left\{f : |f| \leq M_f(1+x^2)\right\},$$

where M_f is positive constant which may depend on f with the norm

$$\|f\|_2 = \sup_{x\geq 0}\frac{|f(x)|}{1+x^2}.$$

$C_2[0,\infty)$ denotes the space of all continuous functions in $B_2[0,\infty)$. For $f\in C_2[0,\infty)$, the weighted modulus of continuity $\Omega(f;\delta)$ is given by

$$\Omega(f;\delta) = \sup_{0\leq\beta<\delta}\frac{|f(x+\beta)-f(x)|}{(1+\beta^2)(1+x^2)}. \tag{3.11}$$

Let

$$C_2^*[0,\infty) := \left\{f\in C_2[0,\infty) : \lim_{x\to\infty}\frac{f(x)}{1+x^2} \text{ exists and finite}\right\}.$$

It is shown in ([12], p. 359, 360) that for every $f\in C_2^*[0,\infty)$, $\Omega(.;\delta)$ has the properties

$$\lim_{\delta\to 0}\Omega(f;\delta)=0,$$

and

$$\Omega(f;d\delta) \leq 2(1+d)(1+\delta^2)\Omega(f;\delta) \qquad p>0. \tag{3.12}$$

For $f\in C_2^*[0,\infty)$, using the properties of Eqs. (3.11) and (3.12), we can write

$$\begin{aligned}|f(t)-f(x)| &\leq (1+(t-x)^2)(1+x^2)\Omega(f;|t-x|) \\ &\leq 2\left(1+\frac{|t-x|}{\delta}\right)(1+\delta^2)\Omega(f;\delta)(1+(t-x)^2)(1+x^2).\end{aligned}$$

Following estimate is the Quantitate direct result for Gupta–Srivastava operators:

Theorem 3.4 *Let $f \in C_2^*[0,\infty)$ such that $f', f'' \in C_2^*[0,\infty)$ and $c \in \mathbb{N} \cup \{0\} \cup \{-1\}$ then for sufficient large n and for each $x \in [0,\infty)$,*

$$\left|n\left(L_{n,c}(f;x) - f(x)\right) - x(1+cx)f''(x)\right| = O(1)\Omega\left(f'', \frac{1}{\sqrt{n}}\right),$$

as $n \to \infty$ holds true.

Proof By Taylor's theorem

$$f(t) = f(x) + f'(x)(t-x) + \frac{f''(x)}{2!}(t-x)^2 + h_2(t,x), \tag{3.13}$$

where

$$h_2(t,x) = \frac{f''(\xi) - f''(x)}{2!}(t-x)^2, \tag{3.14}$$

where ξ lies between t and x.
From the well-known property of the weighted modulus of continuity, we get

$$\left|f''(\xi) - f''(x)\right| \le 4(1+x^2)(1+\delta^2)^2\left(1 + \frac{(t-x)^4}{\delta^4}\right)\Omega\left(f'', \delta\right). \tag{3.15}$$

From (3.14) and (3.15), we get

$$h_2(t,x) \le 8(1+x^2)\Omega\left(f'', \delta\right)\left(1 + \frac{(t-x)^4}{\delta^4}\right)(t-x)^2, \tag{3.16}$$

for $0 < \delta < 1$.
Applying $L_{n,c}(;x)$ on both sides of (3.13) and using Remark 2.1, we get

$$\left|L_{n,c}(f;x) - f(x) - \frac{f''(x)}{2!}L_{n,c}\left((t-x)^2;x\right)\right| \le L_{n,c}\left(|h_2(t,x)|;x\right). \tag{3.17}$$

For the operators [17], the order of convergence is

$$L_{n,c}((t-x)^s;x) = O\left(n^{-\left[\frac{s+1}{2}\right]}\right). \tag{3.18}$$

From (3.16), (3.18) and Lemma 2.2, we get

$$\begin{aligned} L_{n,c}\left(|h_2(t,x)|;x\right) &\le 8(1+x^2)\Omega\left(f'';\delta\right)L_{n,c}\left((t-x)^2 + \frac{(t-x)^6}{\delta^4};x\right) \\ &= 8(1+x^2)\Omega\left(f'';\delta\right)\left(L_{n,c}\left((t-x)^2;x\right) + \frac{1}{\delta^4}L_{n,c}\left((t-x)^6;x\right)\right) \\ &= 8(1+x^2)\Omega\left(f'';\delta\right)\left(O\left(\frac{1}{n}\right) + \frac{1}{\delta^4}O\left(\frac{1}{n^3}\right)\right) \text{ as } n \to \infty. \end{aligned}$$

By Choosing $\delta = \frac{1}{\sqrt{n}}$, we get

$$L_{n,c}\left(|h_2(t,x)|;x\right) = O(1)\Omega\left(f'', \frac{1}{\sqrt{n}}\right). \tag{3.19}$$

From (3.17), (3.19) and Lemma 2.2, we get

$$\left|n\left(L_{n,c}(f;x) - f(x)\right) - x(1+cx)f''(x)\right| = O(1)\Omega\left(f'', \frac{1}{\sqrt{n}}\right),$$

as $n \to \infty$. □

4 Function of bounded variation

The operators (1.1) can be written in the following form:

$$L_{n,c}(f(t);x) = \int_0^\infty U_{n,c}(x;t)f(t)dt,$$

where

$$U_{n,c}(x;t) = \{n + (m+1)c\}\sum_{k=1}^{\infty} p_{n+mc,k}(x;c)p_{n+(m+2)c,k-1}(t;c) + p_{n+mc,0}(x;c)\delta(t),$$

$\delta(t)$ is dirac-delta function.

Lemma 4.1 *For fixed $x \in [0,\infty)$ and if n is sufficient large then, we have*

(i) $\chi_{n,c}(x;y) = \int_0^y U_{n,c}(x;t)dt \le 2\frac{\psi_n^2(x)}{(x-y)^2}, \quad 0 \le y \le x;$

(ii) $1 - \chi_{n,c}(x;y) = \int_z^0 U_{n,c}(x;t)dt \le 2\frac{\psi_n^2(x)}{n(z-x)^2}, \quad x < z < \infty.$

In this section, we study the rate of convergence of the operators (1.1) in the class $DBV[0,\infty)$, and the class of all absolutely continuous functions f defined on $[0,\infty)$ having a derivative coinciding a.e. with a function of bounded variation on $[0,\infty)$. It can be observed that for $f \in DBV[0,\infty)$, we can write

$$f(x) = \int_0^x g(t)\,dt + f(0),$$

where $g(t)$ is a function of bounded variation on each finite subinterval of $[0,\infty)$.

Theorem 4.1 *Let $f \in DBV(0,\infty)$ then for all $x \in (0,\infty)$ and sufficiently large n, we have*

$$\begin{aligned} &\left|L_{n,c}(f;x) - f(x)\right| \\ &\le \sqrt{\frac{1}{2n}}\psi_n(x)\left|f'(x+) - f'(x-)\right| + \frac{x}{\sqrt{n}}\vee_{x-\frac{x}{\sqrt{n}}}^{x}(f_x') \\ &\quad + \frac{2\psi_n^2(x)}{nx}\sum_{k=1}^{[\sqrt{n}]}\vee_{x-\frac{x}{k}}^{x}(f_x') + \frac{2\psi_n^2(x)}{nx^2}\left|f(2x) - f(x) - xf'(x+)\right| \\ &\quad + \frac{2\psi_n^2(x)}{nx}\sum_{k=1}^{[\sqrt{n}]}\vee_{x}^{x+\frac{x}{k}}(f_x') + \frac{x}{\sqrt{n}}\upsilon_x^{x+\frac{x}{\sqrt{n}}}(f_x') + M(r,\alpha,x) \\ &\quad + 2\frac{|f(x)|}{nx^2}\psi_n^2(x) + \sqrt{\frac{2}{n}}\psi_n(x)f'(x+), \end{aligned}$$

where $\vee_a^b f(x)$ *denotes the total variation of* f *on* $[a,b]$ *and* f_x *is an auxiliary operator given by*

$$f_x(t) = \begin{cases} f(t) - f(x-), & 0 \le t < x \\ 0, & t = x \\ f(t) - f(x+), & x < t < \infty. \end{cases}$$

Proof Using Lemma 4.1, the proof of this theorem closely follows the idea which is developed in [14], so we skip the proof. □

5 Weighted approximation

Since the uniform norm is not valid to estimate the rate of convergence in the case of unbounded function defined on the non-compact interval $[0,\infty)$, in this section we study the approximation properties of the operators (1.1) in the weighted spaces of continuous and boundless functions defined on the internal $[0,\infty)$.

Theorem 5.1 *For each* $f \in C_2^*[0,\infty)$*, we have*

$$\left|L_{n,c}(f(t);x) - f(x)\right| \le 4M_f(1+x^2)\delta_n^2(x) + 2\omega_{b+1}(f;\delta_n(x)),$$

where $\delta_n^2(x) = L_{n,c}((t-x)^2;x)$.

Proof From [8], for all $x \in [0,b]$ and $t \ge 0$, we have

$$|f(t) - f(x)| \le 4M_f(1+x^2)(t-x)^2 + \left(1 + \frac{|t-x|}{\delta}\right)\omega_{b+1}(f;\delta_n(x)),\ \delta > 0.$$

Applying Cauchy Schwarz inequality, we get

$$\begin{aligned}\left|L_{n,c}(f(t);x) - f(x)\right| &\le 4M_f(1+x^2)L_{n,c}((t-x)^2;x) \\ &\quad + \left(1 + \frac{L_{n,c}\,|t-x|}{\delta}\right)\omega_{b+1}(f;\delta_n(x)) \\ &\le 4M_f(1+x^2)\delta_n^2(x) + \left(1 + \frac{\delta_n(x)}{\delta}\right)\omega_{b+1}(f;\delta_n(x)).\end{aligned}$$

Now choosing $\delta = \delta_n(x)$, we get the required results. □

Theorem 5.2 *Let* $f \in C_2^*[0,\infty)$*, then, we have*

$$\lim_{n\to\infty} \left\|L_{n,c}(f(t);x) - f(x)\right\|_2 = 0.$$

Proof From [7], it suffices to prove that

$$\lim_{n\to\infty} \left\|L_{n,c}(t^r;x) - x^r\right\|_2 = 0, \quad r = 0,1,2. \tag{5.1}$$

From lemma 2.1, the condition (5.1) is true for $r = 0, 1$.
Now for $r = 2$, we have

$$\begin{aligned}\left\|L_{n,c}(t^2;x) - x^2\right\|_2 &= \underset{x\ge 0}{Sup}\left(\frac{1}{1+x^2}\right)\cdot\left|\frac{(n+(m+1)c)}{(n+(m-1)c)}x^2 + \frac{2x}{(n+(m-1)c)} - x^2\right| \\ &\le \underset{x\ge 0}{Sup}\left(\frac{x^2}{1+x^2}\right)\cdot\left|\frac{2c}{n+(m-1)c}\right| + \underset{x\ge 0}{Sup}\left(\frac{x}{1+x^2}\right)\cdot\left|\frac{2}{n+(m-1)c}\right|,\end{aligned}$$

which implies that

$$\lim_{n\to\infty} \left\|L_{n,c}(t^2;x) - x^2\right\|_2 = 0.$$

Hence the proof is completed. □

Theorem 5.3 *For each* $f \in C_2^*[0,\infty)$ *and* $a > 0$*, we have*

$$\lim_{n\to\infty} \sup_{x\in[0,\infty)} \frac{|L_{n,c}(f;x) - f(x)|}{(1+x^2)^{1+a}} = 0.$$

Proof Let $x_0 \in [0,\infty)$ be arbitrary and fixed. Then,

$$\begin{aligned}\sup_{x\in[0,\infty)} \frac{|L_{n,c}(f;x) - f(x)|}{(1+x^2)^{1+a}} &\le \sup_{x\le x_0} \frac{|L_{n,c}(f;x) - f(x)|}{(1+x^2)^{1+a}} + \sup_{x>x_0} \frac{|L_{n,c}(f;x) - f(x)|}{(1+x^2)^{1+a}} \\ &\le ||L_{n,c}(f) - f||_{C[0,x_0]} + ||f||_2 \sup_{x>x_0} \frac{|L_{n,c}(1+t^2;x)|}{(1+x^2)^{1+a}} + \sup_{x>x_0} \frac{|f(x)|}{(1+x^2)^{1+a}}.\end{aligned}$$

Since $|f(x)| \le ||f||_2(1+x^2)$, we have $\sup_{x>x_0} \frac{|f(x)|}{(1+x^2)^{1+a}} \le \frac{||f||_2}{(1+x_0^2)^a}$.

Let $\varepsilon > 0$ be arbitrary. We can choose x_0 to be sufficiently large that

$$\frac{||f||_2}{(1+x_0^2)^a} < \frac{\varepsilon}{6}. \tag{5.2}$$

Since $\lim_{n\to\infty} \sup_{x>x_0} \frac{L_{n,c}(1+t^2;x)}{1+x^2} = 1$, it follows that

$$\sup_{x>x_0} \frac{L_{n,c}(1+t^2;x)}{1+x^2} \le \frac{(1+x_0^2)^a}{\|f\|_2}\frac{\varepsilon}{3} + 1,$$

for sufficiently large n. Therefore,

$$\|f\|_2 \sup_{x>x_0} \frac{L_{n,c}(1+t^2;x)}{(1+x^2)^{a+1}} \le \frac{\|f\|_2}{(1+x_0^2)^a} \sup_{x>x_0} \frac{L_{n,c}(1+t^2;x)}{(1+x^2)} \le \frac{\varepsilon}{3} + \frac{||f||_2}{(1+x_0^2)^a}.$$

Using Theorem 5.1, we can find for sufficiently large n

$$||L_{n,c}(f;x) - f(x)||_{C[0,x_0]} < \frac{\varepsilon}{3}. \tag{5.3}$$

Combining the relations (5.2) and (5.3), we obtain

$$\sup_{x\in[0,\infty)} \frac{\left|L_{n,c}(f;x) - f(x)\right|}{(1+x^2)^{1+a}} < \varepsilon.$$

□

Acknowledgements The authors are thankful to the referee for his valuable comments leading to the overall improvements in the paper.

References

1. Acu, A.M., Gupta, V.: Direct results for certain summation-integral type Baskakov-Szász operators. Results Math. **72**(3), 1161–1180 (2017)
2. Acu, A.M., Agrawal, P.N., Neer, T.: Approximation properties of the modified Stancu operators. Numer. Funct. Anal. Optim. **38**(3), 279–292 (2017)
3. Bustamante, J.: Bernstein Operators and Their Properties. Birkhäuser, Basel (2017)
4. Deo, N.: Faster rate of convergence on Srivastava-Gupta operators. Appl. Math. Comput. **218**(21), 10486–10491 (2012)
5. Deo, N., Dhamija, M.: Generalized positive linear operators based on PED and IPED. Iran J Sci Technol Trans Sci (2018). https://doi.org/10.1007/s40995-017-0477-5
6. DeVore, R.A., Lorentz, G.G.: Constructive Approximation. Springer, Berlin (1993)
7. Gadjiev, A.D., On, P.P.: Korovkin type theorems. Math. Zametki. **205**, 781–786 (1976)
8. Ibikli, E., Gadjieva, E.A.: The order of approximation of some unbounded function by the sequence of positive linear operators. Turkish J. Math. **193**, 331–337 (1995)
9. Gupta, V., Agarwal, R.P.: Convergence Estimates in Approximation Theory. Springer, Heidelberg (2016)
10. Gupta, V.: An estimate on the convergence of Baskakov–Bézier operators. J. Math. Anal. Appl. **312**(1), 280–288 (2005)
11. Gupta, V., Rassias, ThM, Agrawal, P.N., Acu, A.M.: Recent Advances in Constructive Approximation Theory. Springer, Cham (2018)
12. Ispir, N.: On modified Baskakov operators on weighted spaces. Turkish J. Math. **25**(3), 355–365 (2001)
13. Ispir, N., Yüksel, I.: On the Bézier variant of Srivastava-Gupta operators. Appl. Math. Notes **5**, 129–137 (2005)
14. Maheswari(Sharma), P.: On modified Srivastava-Gupta operators. Filomat **29**(6), 1173–1177 (2015)
15. Neer, T., Ispir, N., Agrawal, P.N.: Bézier variant of modified Srivastava-Gupta operators. Revista de la Unión Matemaťica Argentina **58**(2), 199–214 (2017)
16. Srivastava, H.M., Gupta, V.: A Certain family of summation-integral type operators. Math. Comput. Model. **37**, 1307–1315 (2003)
17. Gupta, V., Srivastava, H.M.: A general family of the Srivastava-Gupta operators preserving linear functions. Eur. J. Pure Appl. Math. **11**(3), 575–579 (2018)
18. Verma, D.K., Agrawal, P.N.: Convergence in simultaneous approximation for Srivastava-Gupta operators. Math. Sci. (2012). https://doi.org/10.1186/2251-7456-6-22
19. Yadav, R.: Approximation by modified Srivastava-Gupta operators. Appl. Math. Comput. **226**, 61–66 (2014)

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

International Journal of Innovations in Scientific Engineering http://www.ijise.in

(IJISE) 2018, Vol. No. 8, Jul-Dec **e-ISSN: 2454-6402, p-ISSN: 2454-812X**

ARTIFICIAL INTELLIGENCE: DEVELOPING SOFTWARE-BASED APPLICATION TO LEVERAGE GENETIC ALGORITHM TO ENHANCE ITS APPLICABILITY IN BUSINESS AND MEDICAL AREAS

Ruchika Goel

Delhi Technological University

ABSTRACT

A method for systems identification using artificial intelligence is presented in this paper. To prove its usefulness applicability, it was verified in the health sector in the clarification of electrophoresis tests, but this application can be implemented in different areas of interest e.g. in business, management, advertising, etc. A wide examination has been done developing the hearty assets of genetic algorithms and proving their abilities across a large range of glitches. We can use this application in many events areas (medical, business, etc.) The resulting references may constitute the basis of future work in the use of genetic algorithms in the business area.

Keywords - artificial intelligence; medical tests; genetic algorithm; systems identification

I. INTRODUCTION

In my research, I developed a software-based application on artificial intelligence principles, more exactly, using genetic algorithms for the identification of specified parameters of a specific curve. The initial segment of paper incorporates a hypothetical introduction of the technique and the second part the representation of trial results in the medicinal field. Albeit these days numerous applications for frameworks recognizable proof are accessible, our change comprises in utilizing hereditary calculations. Broad research has been performed by misusing the powerful properties of hereditary calculations and exhibiting their abilities over an extensive scope of issues. These strategies have picked up acknowledgment as general critical thinking procedures in numerous applications.

II. THE STRUCTURE OF THE CHOSEN GENETIC ALGORITHM

The structure of the chosen genetic algorithm is presented below [1]:

P1: The initial population is created P(t).

P2: The assessment work is connected to every chromosome of the P(t) populace, deciding their wellness esteem.

P3: The population's chromosomes are arranged in light of their wellness esteem given in the past advance.

P4: The best chromosomes are selected, and these will be placed unconditionally in the next population P(t+1).

P5: The chromosomes that are the question the hybrid administrator is then chosen. Their number is indicated by the p_combine esteem, which is, really, the hybrid rate.

P6: The relatives from the past advance are liable to the transformation administrator, coming about new individuals for the P(t+1) populace.

P7: The population P(t+1) is finished with people chosen arbitrarily from the P(t) population.

The steps P3 – P7 are repeated until the specified number of generations is reached.

The chromosome represents the real coding, as rows of real numbers, each gene signifying the output of a coefficient in the specific function. This purpose can be of the polynomial type, or of the type given by the second order system response. For the polynomial function, the value of the first gene will correspond to the highest ranked coefficient and the last to the free term. For the second type, the function's format is as follows [1]:

The chromosome has parameters as well to the value of the reference step used to present the system's response.

This is one of the most important elements of the genetic algorithm because contributes decisively to the convergence of the algorithm. For the given application, the input data are pairs (xi,yi), where i=1...n. These pairs are the coordinates of the points from the given graphical representation, the number of these points were also at the user's choice.

Let dim_pop be the population's dimension and nr_gene be the number of genes a chromosome has.

Compute the fitness value:

III. THE APPLICATION DESCRIPTION

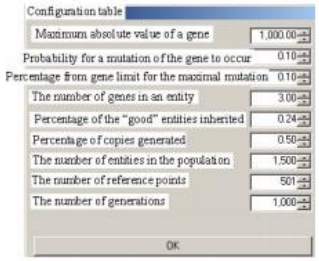
A. The configuration of the genetic algorithm

In the graphical interfaces, the user may set the genetic algorithms configuration, visualize the way the original curve overlaps the computed curve. The computed test parameters are also presented in the graphical interface.

The purpose of these experiments is to detect the application performance in achieving the identification, but also to analyze the influence of different elements of the genetic algorithm on the process of obtaining these results [1].

We configured the genetic algorithm using the values presented in figure 1.

Figure 1. The values of parameters

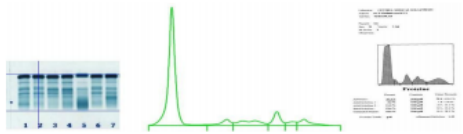


The strong point of our experiment is the identification of the curve parameters followed by the drawing of the curve using the computed parameters. we observe that the computed response (in red) is very close to the original response (in blue). The computed error is under 10%.

B. The application of the genetic algorithm for electrophoresis tests

Usually, an electrophoresis test involves these way [2]:

Figure 2. An example of test's interpretations (the usual case)



The classical results of an electrophoresis test are like five curves. Therefore, in this case, the physician's experience is very important for the good interpretation of results. [2] The main goal of our application is to replace the specialist interpretation when an experienced physician is not available.

In the next, we present the result obtained using the application for this situation.

Figure 3. An example of the test's interpretations using our application.



The curve extracted from electrophoresis gel represents the input data for the genetic algorithm. After the algorithm configuration, in the main window of the application, the computed curves are displayed approximating the original ones. The purpose of our method is the automatic drawing of the five curves. If the curve is drawn exactly, the final step of the test is the computing of maximum values. These values are the values for albumin, gamma globulin, alpha globulin, etc.

The original curve is the representation of normal parameters. Figure 4 presents the superposition of the original curve and computed curves in different moments of the application running.

Figure 4. The preliminary results

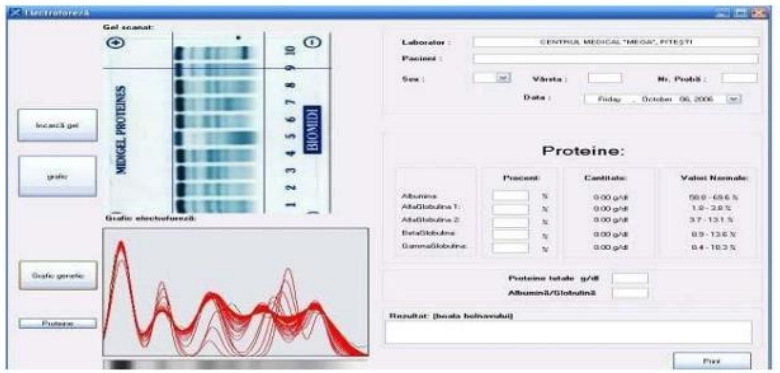
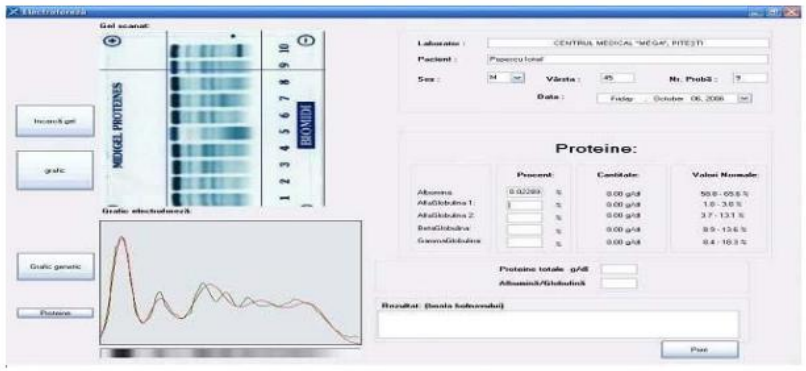


Figure 5. The final results



In figure 5, the computed parameters: albumin, alpha globulin, etc are presented.

IV. FUTURE WORK

Based on our research, genetic algorithms can be implemented in any field besides the medical one, e.g business, logistics, and management. This is possible as long as you can convert it as an optimization problem with a model represented by a list of parameters, which is enabled with genetic operations like mutation, crossover, and selection.

When you have a model, which is ready for optimization using a genetic algorithm, but your design or business intuition is more accurate than the result given by simulation. [4]. Today, genetic algorithms have become the most valuable asset in advertising. Currently, websites are gathering data about their users. Not only that users read but what they read and also verify if they read it in their entirety and how long it took them. On this basis, you can appropriate whether they were interested. After that, all users can be grouped and drawn conclusions. Thanks to genetic algorithms, it is possible to find twins. If 100 of the people who have used the service 40 have common characteristics, it means that when other people with these characteristics are found in the network, the probability of using the service is higher [5].

Another interesting area of research concerning the use of genetic algorithms is scheduling and planning a business process. Scheduling issues where you have to allocate resources to tasks collection are usually difficult to solve due to the presence of multiple constraints [6].

V. CONCLUSIONS

Our study indicates that the genetic algorithms can be successfully used and are more robust than other methods of controlled search; they use performance functions obtained through simple transformations of the function under evaluation. Also, genetic algorithms are easy to use and can find optimal solutions very easily.

This application is an elective strategy for assessment of the lab tests (in exceptional electrophoresis tests), utilizing a piece of artificial reasoning. The primary favorable position of this technique is the requirement for negligible information in the restorative field. Consequently, the execution of this methodology gives an instrument simple to utilize, which offers the tests results rapidly and plainly.

Although the results of research and tests have demonstrated the effectiveness of the use of genetic algorithms, they could be further developed in a number of ways by testing them in many other fields such as business, management, logistics etc.

International Journal of Innovations in Scientific Engineering http://www.ijise.in

(IJISE) 2018, Vol. No. 8, Jul-Dec **e-ISSN: 2454-6402, p-ISSN: 2454-812X**

REFERENCES

[1] C. Săvulescu, I. Sima, E. Sofron, “The functional identification of systems using genetic algorithms”, Proceedings of the 1st International Symposium on Electrical and Electronics Engineering, pp. 82-85, ISBN 973-627-355-3, Galaţi 12-13 October 2006J.

[2] A.T.Andrews, – Theory, Techniques and Biotechnical and Clinical Applications, Oxford University Press, 1993. [3] Al. Ene, C. Savulescu, “Sisteme cu inteligenta artificiala”, Indrumar de laborator, Pitesti, 1998.

[3] [Online] Available: https://www.quora.com/How-can-I-run-simulations-using-genetic-algorithms-in-non-science-fields-e- g-design-or-business, access date 2017-04-09.

[4] [Online] Available: http://www.rp.pl/Biznes-IT/309209922-Sztuczna-inteligencja-przyszloscia-nowoczesnego- biznesu.html#ap-1, access date 2017-04-07.

[5] [Online] Available: http://zeszyty- naukowe.wwsi.edu.pl zeszyty/zeszyt1/Algorytmy_Ewolucyjne_I_Ich_Zastosowania. pdf, access date 2017-04-05.

Automated System for Prediction of Skin Disease using Image Processing and Machine Learning

Lakshay Bajaj
Department of Biotechnology,
Delhi Technological University,
Shahbad Daulatpur, Main
Bawana Road, Delhi-110042,
India

Himanshu Kumar
Department of Biotechnology,
Delhi Technological University,
Shahbad Daulatpur, Main
Bawana Road, Delhi-110042,
India

Yasha Hasija
Department of Biotechnology,
Delhi Technological University,
Shahbad Daulatpur, Main
Bawana Road, Delhi-110042,
India

ABSTRACT
Dermatology is one of the major field of prescription which is worried about the analysis and treatment of skin disorders. Skin diseases are among the most widely recognized medical issues around the world. Regardless of being common, their determination is quite troublesome and requires broad knowledge and expertise in the area. Skin disease might cause severe health and monetary consequences for patients if not detected and controlled early. Early recognition can forestall the condition from worsening. This research paper presents the development of an automated skin disease diagnosis system which takes images of a skin disease as an input by the user and predicts the type of skin disease. The system uses a dual stage approach for detection and prediction process which effectively amalgamates image processing and machine learning. In the 1st stage, the image of the skin condition is subject to numerous types of pre-processing techniques followed by feature extraction. The extracted features for each image are then converted to a feature vector. In the second stage, the feature vectors are fed to a machine learning algorithm (artificial neural networks) to identify disease and predict accordingly. On training and testing for 5 diseases (eczema, psoriasis, impetigo, melanoma, and scleroderma) system produces an overall prediction accuracy of 90%.

General Terms
Dermatology, image processing, machine learning

Keywords
Skin diseases, pre-processing techniques, artificial neural networks

1. INTRODUCTION
Medical Imaging is now extensively used in medical science so that internal structures of the human body can be determined and diagnosed. As it helps to look inside in the body without actually having to open the body [1]. Hence, medical imaging undoubtedly has become an important tool for the detection and evaluation of the treatment of a disease. Skin detection is one of the elementary subjects in image processing. In many instances, example human detection and tracking, visual recognition and face identification, a skin detection system are desirable/required. The concept of skin detection lays foundation for skin diseases detection. Towards this direction, several techniques and approaches have been introduced as of now. In a large proportion of the proposed approaches, researchers have attempted to classify and isolate a feature vector for each and every pixel of image, wherein most of the techniques make use of color spaces to extract required features [2]. In most developing countries, prices are too steep for majority of the population to consult a dermatologist. For example, in a developing country like India, most of the people are reluctant to go to a dermatologist for their skin disease problem as they are not able to afford the prices. Every year significantly great number of people in the developing nations suffer due to diverse range of skin diseases. Therefore, the widespread of skin disease has made it is an indispensable need for both the patients and dermatologists to have an automated skin disease detection system especially in developing countries [3].Our application involves a layered approach, with two stages: first, we make use of Image Processing for identification as well as extraction of region of interest and the second stage utilizes Machine Learning for a fairly accurate prediction of the disease. This system is capable of detecting 5 main skin disorders namely psoriasis, melanoma, scleroderma, eczema, impetigo.

2. RECENT WORKS
In the past two decades, strong impulse has been given to developing automated systems capable of assisting physicians in medical imaging task [4].The paper proposed by Muhammad Zubair Asghar et al [5] presents an online expert system to detect certain skin diseases that make use of forward-chaining with depth-first search method. However, using a system based on set rules and symptoms is not feasible due to the various manifestations of a single skin disease. The system created by us would be superior in performance in such a manner as the issue we are addressing is probabilistic in itself and thus we require a system which has the ability to adapt and learn from the underlying pattern that exists in the skin disease which can be inferred by the image. Another system that was proposed by A.A.L.C. Amarathunga et al [6] introduces a data mining unit to the system of skin detection. However, it lacks the choice of attributes essential for detection. Neither the data source nor the attributes used for learning/testing have been mentioned. M. SHAMSUL et al [7] proposed a system that uses an image processing system with pre-processing algorithms and a feed forward neural network**.** Florence et al [8] categorize the image as a bacterial or viralskin infection using image processing techniques. Damilola et al [9] developed a system that assembles pigmented skin lesions image results, analysis, comparing observation and conclusions by medical experts using prototyping methodology. Our system makes use of image processing with pre-processing algorithms and feed forward back propagation method in artificial neural networks that are discussed in the following section.

3. ARCHITECTURE AND METHODOLOGY
The system amalgamates two approaches that are image processing and machine learning algorithms represented in figure 1.

The first stage makes use of different image processing unit to isolate and extract the region of interest that is the diseased region from the image.

Initially those pre-processing algorithms were applied on colour skin images and further image processing algorithms were applied. The Second stage makes use of machine learning unit, the input images are fed to an artificial neural network to train network and predict an accurate result. For all the process to take place we needed data in the form of skin disease image, particularly of eczema, psoriasis, impetigo, melanoma, and scleroderma. The training data set was obtained from [10].

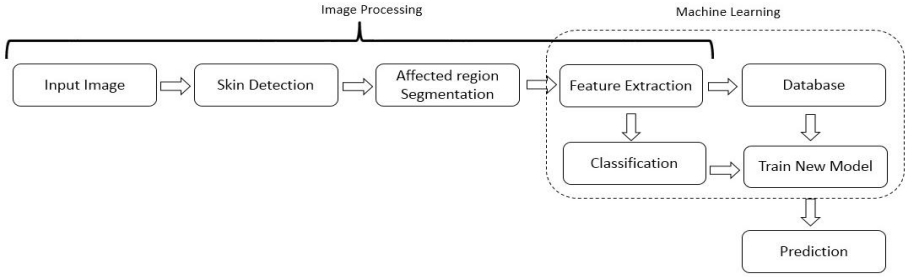


Fig 1: Brief Flowchart of the methodology used

3.1 Image Processing Unit
The colour images taken from the database were subjected to pre-processing algorithms before feature extraction. However, there are various image processing methods available to extract the required region such as thresholding, compression based methods, clustering, region growing, watershed transformation, etc. From previous works, we found that the edge-based segmentation method based on active contour, can be used to segment the affected region more accurately [11]. The disease region was segmented using the active contour and pre-processing algorithm. Both the methods were compared and gave equal results but it was found that active contour had a flaw, the user needed to mark the region around the infected area. Therefore, Pre-Processing Algorithm was given preference over active contour because of its autonomous nature and reducing the involvement of user.

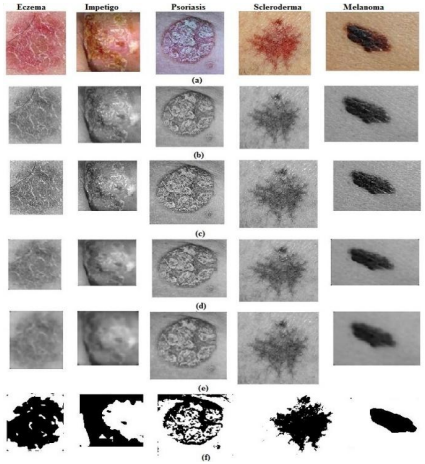


Fig 2: Result of Pre-processing algorithms. (a) Original Color skin image, (b) Grey Scale image, (c) sharpening filter, (d) Median filter, (e) Smooth Filter, (f) Binary Mask

3.2 Feature Extraction
In order to improve the accuracy of feature extraction pre-processing algorithms were used. The algorithms used were grey scale, sharpening filter, median filter, smooth filter, binary filter, and RGB extraction. RGB values of the image were stored beforehand converting them to grey scale image. Sharpening filter was then applied to enhance the subtle details of the grey scale image. Median filter was then used to remove any noise from the image using a 5*5 matrix. The next filter used was the mean filter or smoothening filter which replaced each and every pixel with average (mean) value of its neighbor including itself. The resulting image was then converted into binary image which provides the basis of extraction of diseased region. The result of the applied filter is represented in figure 2. The Sobel operator was then used to detect edges which facilitated us in detecting the shape of the affected region. The affected/diseased region was extracted and compared to RGB color stored before converting to grey scale which resulted in extraction of only the colored diseased region of the image.

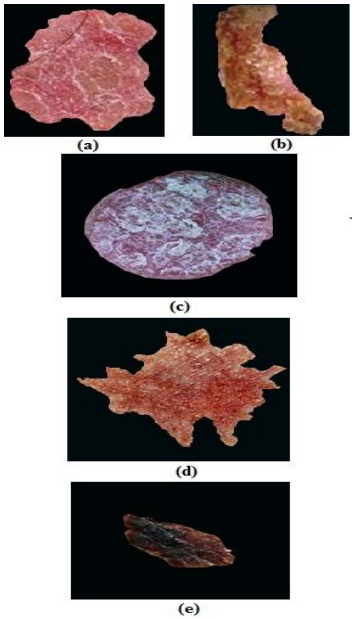


Fig 3: Segmented Disease region (a) Eczema (b) Impetigo (c) Psoriasis (d) Scleroderma (e) Melanoma

This was done by super-imposing the binary image and the original color image pixels and then elimination of those pixels which were not present in the binary affected region. Figure 3 represents the resulting images.

The resulting diseased region was then converted into a feature vector. Similarly, each image in the training set was converted into feature vector in MATLAB and classified together so that it can be then fed to the machine learning unit for training and testing. Figure 4 represents the feature vector obtained from each image. Each feature vector contains 25 feature values from each image in column-wise orientation.

3.3 Machine Learning Unit

We used feed forward back propagation artificial neural network (ANN) for this step. The ANN comprised of one input layer, two hidden layers and one output layer. Since, we have used a supervised machine learning algorithm, the feature table obtained after image processing was fed to the input layer and their corresponding results to the output layer for training. A deep learning technique using the neural network tool in MATLAB which trains the model using an identity function was used to initialize the parameters of the model, the weight matrices W.

• Input Layer: Input from the user (25 neurons)

• Hidden Layer: Composed of sigmoid function (40 neurons)

• Output Layer: Softmax layer

Once the input and output data were fed the network was trained. Further, the system was validated and tested by using the tenfold cross validation process because of its unique quality which makes sure that there is no overlapping of the training data and the test data, thereby, making the test results fairly accurate and dependable. The system was trained using 25 features as input and made use of 40 neurons in the hidden layer. Figure 5 represents how our system is trained by using 25 features from each image. Output layer represents 5 neurons because we used a softmax layer [12] approach as it gives probabilistic results for all 5 diseases.

FEATURE TABLE

	IMG 1	IMG 2	IMG 3	IMG 4	IMG 5	IMG 6	IMG 7	IMG 8
1	222	190	211	150	210	175	128	174
2	216	187	209	156	203	181	135	168
3	227	188	211	145	203	185	140	168
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
24	208	207	145	203	171	211	204	173
25	195	193	173	209	169	215	208	171

Fig 4: Feature Table

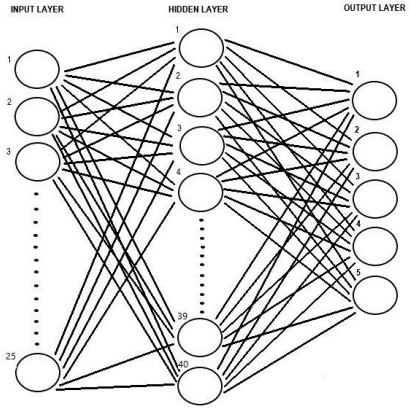


Fig 5: Feed Forward Back Propagation Neural Network

In **forward propagation**, a set of weights are applied to the input data for the calculation of an output. Weights are selected randomly for the first forward propagation.

In **back propagation**, margin of error is measured for output and accordingly, weights are adjusted to decrease the error.

4. RESULTS AND DISCUSSION

The approach of using 2 stage process for prediction of skin diseases, where the disease region is converted into a feature vector and then used for training of network is novel to best of our knowledge with an overall accuracy of 90%.Correlation of our work with related works in this area has uncovered stark contrasts in the execution and performance. None of the current solutions for detection and prediction of skin diseases deals with the five skin diseases that we have proposed in approach. . Out of the 813 images of 5 diseases, 70% of the images were used for training, 15% for validation and

remaining 15% for testing. Figure 5 represents regression graph and their respective accuracy rate.

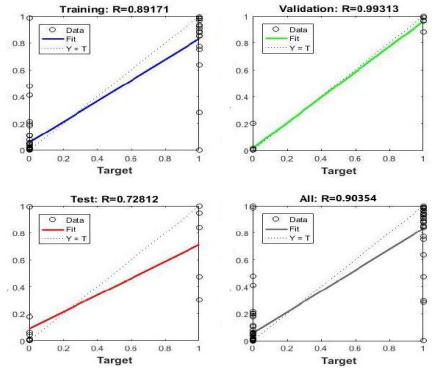


Fig 6: Regression graph representing training, validation, test and overall accuracy rate

Out of 5 skin diseases used, the detection rate of Eczema and Psoriasis was 92.5% and 91.6% respectively. One of the proposed methods presented in [3] obtained only 88% detection accuracy for Eczema. Similarly, detection rate of 88% and 89.3% for Eczema and Psoriasis respectively could be obtained in [9] which is less than the detection rate obtained by our system. Although, detection rate of Scleroderma was not up to our expectations, however, to best of our knowledge, no system as of yet has been proposed to successfully detect Scleroderma. Figure 6 shows individual detection rate of diseases.

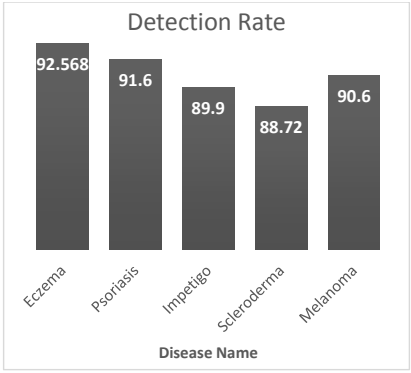


Fig 7: Detection Rate of Each Disease

5. ACKNOWLEDGMENTS

Thanks to all the experts and Department of Biotechnology, Government of India [No.BT/PR5402/BID/7/408/2012] for funding this project.

6. REFERENCES

[1] Aarthipoornima E, Jeyaseelan T, Medical imaging modalities: A survey, in: 2016 Int. Conf. Emerg. Trends Eng. Technol. Sci., IEEE, 2016: pp. 1–4. doi:10.1109/ICETETS.2016.7603066.

[2] Shervan F. E, Mohammad S, Farshad T, An Innovative Skin Detection Approach Using Color Based Image Retrieval Technique, Int. J. Multimed. Its Appl. 4 (2012) 9. doi:10.5121/ijma.2012.4305.

[3] Rahat Y, Md. Ashiqur R, Nova A, Dermatological disease detection using image processing and artificial neural network, in: 8th Int. Conf. Electr. Comput. Eng., 2014: pp. 687–690. doi:10.1109/ICECE.2014.7026918.

[4] Radu D, Stefan M, Dan P, Medical images classification for skin cancer diagnosis based on combined texture and fractal analysis. WSEAS Transactions on Biology and Biomedicine, July 2010.

[5] Muhammad Z. A, Asghar Mj, Sheikh S, Shakeel A, Diagnosis of Skin Diseases using Online Expert System, International Journal of Computer Science and Information Security, June 2011

[6] A.A.L.C. Amarathunga, E.P.W.C. Ellawala, G.N. Abeysekara, C.R.J. Amalraj, Expert System For Diagnosis Of Skin Diseases, Int. J. Sci. Technol. Res. 4 (2015) 174–178.

[7] M. Shamsul Arifin, M. Golam Kibria, Adnan. F, M. Ashraful Amini, Hong Y, Dermatological disease diagnosis using color-skin images, Published in Proceedings of the 2012 International Conference on Machine Learning and Cybernetics, Xian, 15-17 July, 2012

[8] Florence T,Ernest M, Fred N. K, An image-based diagnosis of virus and bacterial skin infections, International Conference on Computing and ICT Research,2011

[9] Damilola A. O, Olidayo O. O, Soloman A. O, Automating skin disease diagnosis using image classification, Published in Proceedings of the World Congress on Engineering and Computer Science 2013 Vol II WCECS 2013, 23-25 October, 2013, San Francisco, USA

[10] Diepgen TL, Yihune G et al. Dermatology Online Atlas. Published online at: http://www.dermis.net/doia/

[11] Teck T. T, Li Z, Ming J, , An intelligent decision support system for skin cancer detection from dermoscopic images in 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)

[12] Er. Shrinidhi G, Ansari N, Ansari Z Shaikh R,An Innovative Approach for Skin Disease Detection Using Image Processing and Data Mining. In International Journal of Innovative Research in Computer and Communication Engineering.

Wireless Personal Communications
https://doi.org/10.1007/s11277-019-06121-7

Adaptive Technique with Cross Correlation for Lowering Signal-to-Noise Ratio Wall in Sensor Networks

Rajiv Kapoor[1] · Rashmi Gupta[2] · Le Hoang Son[3,4] · Sudan Jha[5] · Raghvendra Kumar[6,7]

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract
Energy detection is considered as most popular spectrum sensing technique. However, the impact of noise uncertainty in wireless environment limits the sensitivity of energy-based detector. It generally suffers from Signal-to-Noise Ratio (SNR) wall, which is defined as the minimum SNR in which robust detection is not possible. This paper presents an adaptive technique to improve SNR wall in sensor networks that combines cross-correlation scheme and dynamic threshold method to improve performance. An analytical expression of SNR wall is derived to show improved performance over traditional energy detector. Theoretical analyses and simulations validate effectiveness of the proposed method under noise uncertainty environment.

Keywords Adaptive threshold · Cognitive radio · Cross-correlation · Power recognition · Noise uncertainty parameters · Signal-to-noise ratio wall

✉ Le Hoang Son
lehoangson@tdtu.edu.vn

Rajiv Kapoor
Rajivkapoor.dtu@gmail.com

Rashmi Gupta
rashmig71@yahoo.com

Sudan Jha
jhasudan@hotmail.com

Raghvendra Kumar
raghvendraagrawal7@gmail.com

[1] Department of Electronics & Communication Engineering, Delhi Technological University, Delhi, India

[2] Department of Electronics & Communication Engineering, AIACT&R, Delhi, India

[3] Division of Data Science, Ton Duc Thang University, Ho Chi Minh City, Vietnam

[4] Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam

[5] School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, India

[6] Computer Science and Engineering Department, LNCT College, Bhopal, MP, India

[7] VNU Information Technology Institute, Vietnam National University, Hanoi, Vietnam

Published online: 29 January 2019

1 Introduction

Power recognition can be considered as popular spectrum sensing technique, though the impact of noise uncertainty in wireless environment limits the sensitivity of energy-based detector and generally suffers from the Signal-to-noise ratio (SNR) walls defined as the smallest SNR, where robust detection is not possible [12, 23–25]. The essential benefit of cognitive radio (CR) is its potential toward applying certified frequency spectrum of the primary users (PUs) [31, 55]. CR has emerged as one of the solutions to enhance spectrum performance [11]. It is mainly consisting of two main roles specifically sensing and throughput optimization. Spectrum sensing is improved by enforcing Peak vs. Average Power Ratio of the acquired signal in strength detection [1]. It has some additional advantages such as prior authentication is not required and loosed from sound ambiguity [49]. CR is not effectively relevant to traditional networks. It has the capacity to enhance the overall performance of wireless sensor networks [6, 10].

A cognitive radio detects unused spectrum band and utilizes it in opportunistic way. Energy detection is the most commonly used method to detect presence of unused band [10]. Signal detection at noise uncertainty situation imposes restriction on sensing called the Signal-to-Noise Ratio (SNR) wall [4, 7, 18]. Various methods were proposed for solving the Signal-to-Noise Ratio (SNR) wall problem [14, 30]. Some typical ones include the oversampling with over sampled alerts, the blind spectrum sensing scenario, the non-linear stochastic resonance, adaptive threshold strategies through the most likelihood estimator or second and fourth order second estimator, and correlation strategies [19, 26, 31, 37, 38].

In this article, we proposed a hybrid method to improve the SNR wall. In the first stage, cross-correlation method is used, and in the second stage the adaptive threshold method is used to improve the performance. The main contribution of this article is to derive a new expression for SNR wall by combining the cross-correlation method and adaptive threshold method. We will analyze the shortcomings of the previous works and the idea of innovation in Sect. 2.

The rest of the articles are prepared into four different sections. Section 2 describes related researches for the SNR Wall. Section 3 affords the formula and new model. Section 4 suggests performance analyses and simulation. Finally, conclusions are made in Sect. 5.

2 Related Works

Various methods had been proposed inside the literature for solving the Signal-to-Noise Ratio (SNR) wall problem. Oversampling with over sampled alerts has high correlation among their successive samples compared to noises changed. Later, Shaikh et al. [37] proposed a blind spectrum sensing situation facilitate. Dutta and Manna [9] proposed a new hybrid frame structure for spectrum sensing which has better overall performance than traditional approaches in throughput [10]. The chance of detection at distinct alarm probabilities was evaluated.

Gunes et al. [15] tested resonator parameters, and Liu et al. [27] investigated phenomenon of stochastic resonance in a mono-threshold-system-primarily based detector with additive heritage noise and multiplicative external noise. Ali et al. [2] described an adaptive threshold method with unfixed and adaptive thresholds toward channel situations to

improve detection performance. It calculates a required threshold as a linear feature of recorded sign's general deviation [5]. Zulfikar et al. [55] used cognitive radio generation to handle the problem of inefficiently gadget radio spectrum and boom array use of spectrum. Detection overall performance of energy is relatively structured upon noises even as SNR is very low for PU.

Ali et al. [1] supplied a comprehensive survey on the cognitive radio community paradigm elegant grid, consisting of the tool structural design, statement community composition, application, as well as cognitive radio-based totally communication technology. Possible programs of cognitive radio-primarily based system and cognitive radio-based totally spectrum sensing tactics with their important classifications have been examined with routing and MAC protocols. Some open research challenges faced by cognitive radio-primarily based smart grid networks along have been highlighted therein.

Another technique based totally on adaptive threshold changed into was shown in [19] with spectrum sensing errors being minimized underneath constraints. This comes from the truth that practical systems have spectrum utilization adjustments with wide variety in energetic transmission with corresponding parameters. Oude et al. [31] offered a correlation technique with energy detection. Besides those works, there are a few relevant works on SNR in [11, 20, 36]; Haijun et al. [17].

Nevertheless, they did not reduce SNR Wall as expected. Spectrum sensing schemes are commonly used in many networks. In this paper, the spectrum sensing schemes are used to deal with noise uncertainty and signal-to-noise ratio (SNR) wall. Until the sensing time is not evaluated, we cannot achieve desired performance through signal detection. Therefore, spectrum sensing schemes are used to explore and improve the spectrum's efficiency. They are used for many purposes like identifying frequency holes, assigning frequencies to the spectrum for analyzing of zero harmful interference to the channel, variations in the noise as a function of real time, and enhancing the low signal-to-noise ratio (SNR).

We will show in the Sect. 3 that without the sensor detector, it is unable to achieve the desired performance. Due to this, the robustness against power noise uncertainty will be unpredictably low. Therefore, this scheme is used for both theoretical and experimental analysis to confirm the SNR wall existence under the noise power uncertainty conditions. The spectrum sensing technique used in this article analyzes performance of strength detector that maybe degraded. The noise uncertainty is indeed integrated into the system. In the cross-correlation method, same signs pass into two unique paths which lessen uncorrelated noises. Therefore, SNRs and SNR wall are progressed at outputs. The new hybrid technique based on cross-correlation and adaptive threshold technique effectively treats the drawbacks of existing works. An expression for SNR wall turned into derived that involve to correlation ε, uncertainty U and the dynamic threshold issue d. Simulations will be done to validate the new derived SNR wall in improving energy detection performance in comparison to correlation based methods.

3 Proposed Method

3.1 Problem Formulation

Spectrum sensing using power recognition can be considered as the binary hypothesis testing problem given as:

(i) H_0: represent that the primary user is absent.
(ii) H_1: represent that the primary user is present.

Let $S(N)$ be transmitted primary user through channel $l$and $h = 0$. The received signal $Y(N)$ is given by,

$$Y(N) = S(N) + W(N), \tag{1}$$

where $W(N)$ be an additive white Gaussian noise. Let $D(Y)$ denotes the decision statistics to decide H_1 or H_0.

$$\begin{aligned} 1D(Y) &= \frac{1}{N}\sum_{n=0}^{n} \mathrm{Y(N)} \\ D(Y) &\geq \gamma \ : H_1 \\ D(Y) &< \gamma \ : H_0 \end{aligned} \tag{2}$$

Noise variance is completely known. From the middle limit theorem, the subsequent approximation is preserved [36].

$$D\big(Y/H_0\big) \sim N\big(\sigma_n^2, 2\sigma_n^4/N\big) \tag{3}$$

$$D\big(Y/H_1\big) \sim N\big(p+\sigma_n^2, (p+2\sigma_n^2)^2/N\big) \tag{4}$$

Different probabilities terms such as possibility of discovery $P_d = \mathrm{prob}\{\mathrm{Decision} = H_1/H_1\}$, false alarm probability $P_f = \mathrm{prob}\{\mathrm{Decision} = H_1/H_0\}$ and probability of miss $P_m = \mathrm{prob}\{\mathrm{Decision} = H_0/H_1\}$ are given in Eqs. (5–6).

$$P_{FA} = P_r\big(D(Y/H_0) > \gamma\big) = Q\left(\frac{\gamma - \sigma_n^2)}{\sqrt{2\sigma_n^4/N}}\right) \tag{5}$$

$$P_D = Q\left(\frac{\gamma - (p+\sigma_n^2)}{\sqrt{2(p+\sigma_n^2)^2/N}}\right) \tag{6}$$

From Eqs. (5) and (6), expression for N can be calculated to achieve desired P_d, by fixing P_{FA} to a certain value and for known SNR value. Quantity of sample essentially is given by Eq. (7).

$$N = 2\big[Q^{-1}(P_{FA}) - Q^{-1}(P_D)\big]^2 \times SNR^{-2} \tag{7}$$

However, the noise influences are predictable with distributional ambiguity in a particular gap:

$$\sigma_n^2 \in \left[\frac{\sigma_n^2}{\mu}, \mu\sigma_n^2\right]$$

where $\mu > 1$ is the noise uncertainty coefficient and $U = 10\log_{10}\mu$. From (7), we obtain

$$N = 2\left[\mu Q^{-1}(P_{FA}) - \left(\frac{1}{\mu} + SNR\right)Q^{-1}(P_D)\right]^2 \times \left(SNR - \left(\mu - \frac{1}{\mu}\right)\right)^{-2}$$
$$N \to \infty, SNR \to \frac{\mu^2 - 1}{\mu^2} \tag{8}$$

The **SNR Wall** in terms noise uncertainty factor is given as,

$$SNR_{Wall} = \frac{\mu^2 - 1}{\mu^2} \tag{9}$$

It is recognized that signal detection SNR_{Wall} is unreliable.

3.2 Proposed Hybrid Method

Figure 1 shows block diagram for the proposed method in which both cross-correlation and dynamic threshold techniques are employed simultaneously to improve SNR Wall. Let us transmit PU signal $S(t)$ distorted with channel noise $n(t)$. Here, $n(t) = n_c(t) + n_I(t)$ where $n_c(t)$ is correlative noise component and $n_I(t)$ is interference noise component ($i = 1$, 2-signal travelling through path 1 and path 2).

The *first part* of the block diagram shows the cross-correlation method, as signal travels through path1 and path2 of receiver cross-correlation will reduce interference noise. The *second part* of the method shows the adaptive threshold based on energy detector which further improves SNR and detects unused primary user signals. Let ε be the noise correlation between two receivers.

Let us denote,

$$psd \text{ of } n_c(t) = \varepsilon N_0$$
$$psd \text{ of } n_1(t) = psd \text{ of } n_2(t) = (1 - \varepsilon)N_0 \tag{10}$$

At the output, SNR $= P_s/N_oW.N_oW$ is the noise within band of interest. From Fig. 1, we clearly see that

$$\mathrm{Y} = Y_{\mathrm{re}} + Y_{\mathrm{im}} \tag{11}$$

where Y_{re} and Y_{im} are the real and imaginary parts of the sample respectively because every samples are independent and identically distributed (IID), after that the central limit theorem for a large number of samples, their common converge in the direction of a standard allocation. Here, we are assuming that both receivers in Fig. 1 are perfectly in phase i.e. $\Delta\emptyset = 0$. Therefore, imaginary parts can be assumed as zero. For simplicity purpose, we are taking $N_o = 1$.

Case 1: Under hypothesis H_0 (PU absent $SNR = 0$), mean and variance are given as,

$$\mu_0 = \varepsilon \text{ and } \sigma_0^2 = \frac{1 + \varepsilon^2}{2N} \tag{12}$$

Case 2: Under hypothesis H_1 (PU is present $SNR > 0$), mean and variance are given as

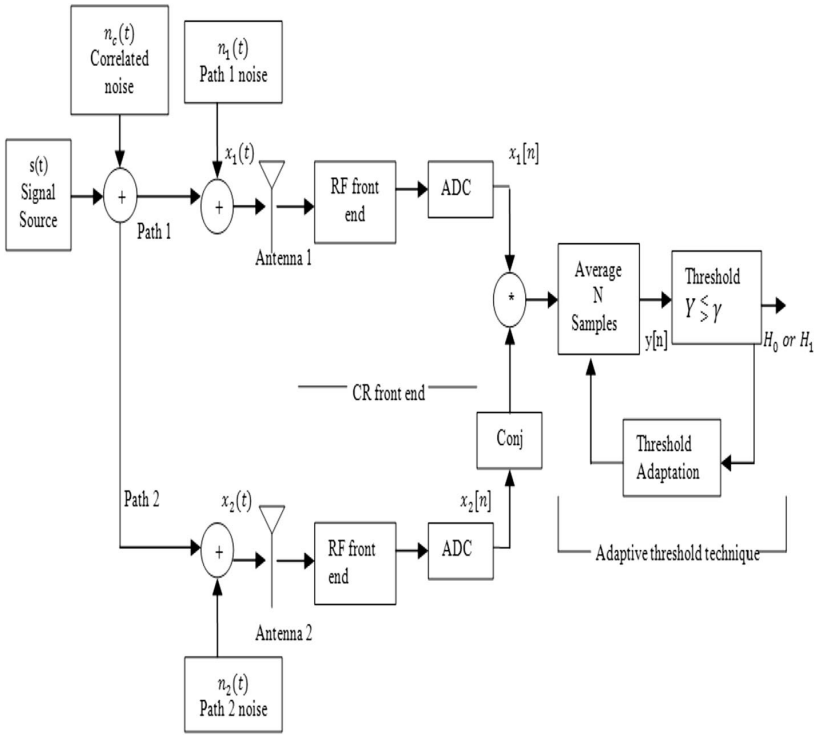


Fig. 1 Architecture of the proposed method

$$\mu_1 = \varepsilon + SNR$$
$$\sigma_1^2 = \frac{2SNR^2 + (2 + 2\varepsilon)SNR + 1 + \varepsilon^2}{2N} \tag{13}$$

P_{FA} and P_D for fixed threshold energy detector method are given as,

$$P_{FA} = Q\left(\frac{\gamma - \mu_0}{\sigma_0}\right) \tag{14}$$

$$P_D = Q\left(\frac{\gamma - \mu_1}{\sigma_1}\right) \tag{15}$$

Due to noise uncertainty for the fixed threshold method, performance of detector also degrades. Assume threshold γ is varied in such a way that system performance is maintained constant to its previous desired value, Here, d is called as dynamic threshold factor ($d > 1$) then (14) and (15) are modified as,

$$P_{FA} = Q\left(\frac{\gamma' - \mu_0}{\sigma_0}\right) = Q\left(\frac{\gamma d - \mu_0}{\sigma_0}\right) \tag{16}$$

$$P_D = Q\left(\frac{\gamma' - \mu_1}{\sigma_1}\right) = Q\left(\frac{\gamma'/_d - \mu_1}{\sigma_1}\right) \tag{17}$$

From Eq. (17), for desired value of P_{FA}, setting the value of γ and multiplying with U factor for the evaluation under the noise uncertainty condition is as below:

$$\gamma = U\frac{[\mu_0 + \sigma_0 Q^{-1}(P_{FA,des})]}{d} \tag{18}$$

Putting value of γ in Eq. (17), we obtain

$$P_D = Q\left(\frac{\frac{[\mu_0+\sigma_0 Q^{-1}(P_{FA,des})]}{d^2} - \mu_1}{\sigma_1}\right) \tag{19}$$

Putting values of $\mu_0,\ \mu_1,\ \sigma_0\ , \sigma_1$ in Eq. (19), we gain

$$P_D \approx Q\left[\frac{\left(\varepsilon\left(U - d^2\right) - d^2 SNR\right)\sqrt{N} + U\sqrt{P}Q^{-1}\left(P_{FA}, des\right)}{d^2\sqrt{\gamma + (\varepsilon+1)SNR}}\right] \tag{20}$$
$$p \equiv 0.5(1+\varepsilon^2),\ SNR \ll 1 \& N \gg 1$$

For $SNR_{\min}$ expression, we approximate Eq. (20) for $N \gg 1$ and given P_D and P_{FA}.

$$SNR_{\min} \approx \frac{\varepsilon\left(U - d^2\right)}{d^2} + \frac{X^2(\varepsilon+1)}{2N} + \frac{UpY}{d^2} - \frac{X\sqrt{Nd^6\left(pd^2 + \left(U + d^2\right)\left(\varepsilon + \varepsilon^2\right)\right) + Z}}{N}$$
$$(P_D \geq P_{D,des}) \,\&\, (P_{FA} \leq P_{FA,des})$$
$$X \equiv Q^{-1}(P_{D,des}),\ Y \equiv Q^{-1}(P_{FA,des})$$
$$Z \equiv X^2(\varepsilon+1)^2 d^8 + pUY(\varepsilon+1)\sqrt{N}d^6 \tag{21}$$

The above expression can be simplified by using $(N \gg 1)$ and $|X|$ and $|Y|$ are fairly small such that $(|X| \cdot |Y| \leq 10)$, which gives $P_{FA} = P_{MD} \geq 10^{-23}$.

$$SNR_{\min} \approx \frac{\varepsilon\left(U - d^2\right)}{d^2} + \frac{UpY - Xd^5\sqrt{pd^2 + \varepsilon\left(U - d^2\right)(1+\varepsilon)}}{d^2\sqrt{N}} \tag{22}$$

Since SNR Wall is defined as a minimum SNR for infinitely long averaging below which a signal cannot be detect. Thus, SNR Wall is expressed as,

$$SNR_{wall_new} = \log_{N\to\infty}\left(SNR_{min}\right) = \frac{\varepsilon\left(U - d^2\right)}{d^2} \tag{23}$$

For the fixed threshold model $(d = 1)$, the above expression becomes

$$snr_{wall_new} = \varepsilon(U - 1) \tag{24}$$

The mathematical expression given in Eq. (23) represents the SNR Wall for proposed method which depends upon three factors:

1. Uncertainty factor (U).
2. Cross-correlation between two paths (ε).
3. Dynamic threshold factor (d).

As the value of U increases the SNR Wall, to compensate the effect of U, the value of ε should maintain low and the value of d should be greater than 1. SNR Wall for $d = 1$, shown in Eq. (24) which represents for fixed threshold energy detection method.

4 Experimental Results

This section describes performance of proposed algorithm to verify the accuracy of the derived relation on SNR degradation and packet loss charge. The SNR_{wall} are verified by the real scenario simulation performed using MATLAB in accordance with the parameters presented in IEEE 802.22 standards. Results obtained from previous techniques [26, 28] are compared with those of the proposed method. Assuming value of P_{MD} and $P_{FA} = 10^{-5}$, performance of SNR_{min} is plotted in Fig. 2 at different noise uncertainty levels. The values of P_{MD} and P_{FA} are 10^{-5}, performance of SNR_{min} are then plotted in Fig. 2 with different noise uncertainty levels.

From the proposed method described in Sect. 3.1, it is clear that the proposed method is employed to improve SNR Wall by using transmitted signal S(t), which is distorted with channel noise n(t). Referring to the Fig. 1, the block diagram possesses two parts. The first part shows the cross-correlation method when signal is traveling through path1 and path2. At the receiver's end, there is cross-correlation which reduces interference noise. Considering the second part of the method, the energy detector improves SNR and detects unused

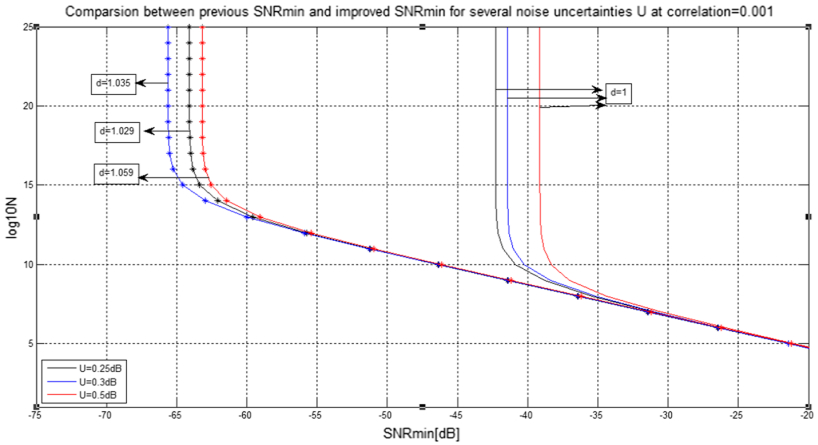


Fig. 2 Comparison of the Previous $SNR_{\min}$ [26, 28] and the Improved $SNR_{\min}$ for several noise uncertainties U and $\varepsilon = 0.001$

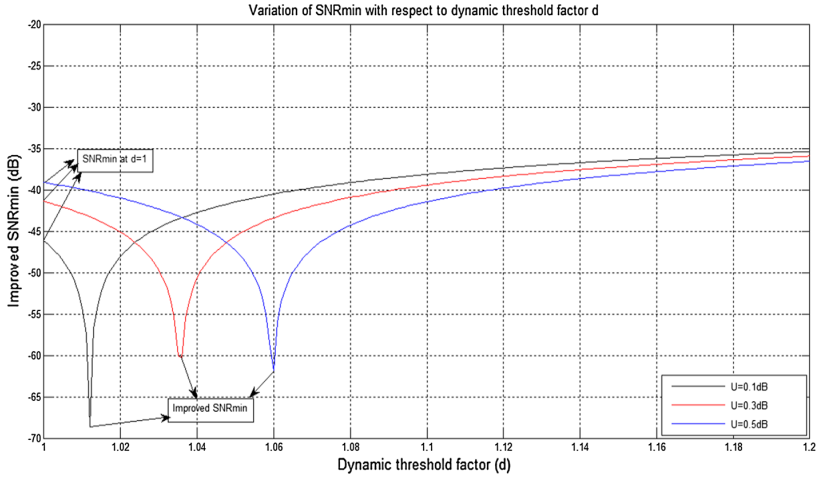


Fig. 3 Variation of $SNR_{\min}$ (dB) with respect to the dynamic threshold (d)

Table 1 Improved $SNR_{\min}$ and percentage of improvement by d at $U=0.1$ dB

No.	$SNR_{\min}$ (d = 1)	Value of 'd'	Improved $SNR_{\min}$	Improvement (%)
1.	−46.327	1.008	−51.478	11.11
2.	−46.327	1.009	−52.907	14.20
3.	−46.327	1.010	−55.044	18.81
4.	−46.327	1.011	-59.404	28.22

primary user signals. Thus, at the output, SNR is the noise within band of interest. From Fig. 1, it can be seen that both are the real and imaginary parts. The reason is because, every samples are independent and identically distributed (IID), after that the central limit theorem for a large number of samples, their common converge is in the direction of a standard allocation. Equations (16–20) indicate the noise uncertainty, which is for a fixed threshold method. However, this degrades performance of detector. This is maintained by the constant in Eqs. (16, 18, 19) in such a way that the system performance is maintained constant to its previous desired value.

Figure 3 shows the $SNR_{\min}$ verses the number of samples at $U=0.25$ dB, 0.3 dB, 0.5 dB and $\varepsilon = 0.001$. The comparison is expressed between $SNR_{\min}$ at $d=1$ and the improved $SNR_{\min}$ for different values of d. We can observe that at $d=1$, $SNR_{min}=-42$ dB is improved to -64 dB for $d=1.029$. Figure 2 shows $SNR_{\min}$ with respect to the dynamic threshold factor 'd'. $SNR_{\min}$ can be obtained from Fig. 2. Clearly, SNR Wall is at -62 dB for $U=0.5$ dB, and $f=1.06$. The variation of SNR_{min} at various values of d and percentages of improvement with $SNR_{\min}$ at $d=1$ is shown in Tables 1 and 2 ($U=0.1$ dB and 0.5 dB respectively). These tables show percentage improvement in $SNR_{\min}$ for correlation noise factor $\varepsilon = 0.001$ and the number of samples $N = 10^{20}$.

Table 2 Improved $SNR_{\min}$ and percentage of improvement by d at $U=0.5$ dB

No.	$SNR_{\min}$ (d = 1)	Value of 'd'	Improved $SNR_{\min}$	Improvement (%)
1	−39.135	1.03	−42.395	8.33
2.	−39.135	1.07	−46.992	20.08
3.	−39.135	1.08	−44.196	12.93
4.	−39.135	1.09	−42.547	8.72

Table 1 shows the $SNR_{\min}$ verses the number of samples at $U=0.1$ dB and $\varepsilon = 0.001$. The comparison is expressed between $SNR_{\min}$ at $d=1$ and the improved $SNR_{\min}$ for different values of d. We can observe that at $d=1$, $SNR_{\min}=-46.327$ dB is improved to -59.404 dB for $d=1.011$ and the percentage of improvement is 28.22%. Table 2 shows the $SNR_{\min}$ verses the number of samples at $U=0.5$ dB and $\varepsilon = 0.001$. The comparison is expressed between $SNR_{\min}$ at $d=1$ and the improved $SNR_{\min}$ for different values of d. We can observe that at $d=1$, $SNR_{\min}=-39.135$ dB is improved to -46.992 dB for $d=1.07$ and the percentage of improvement is 20.08%.

Figure 4 shows the probability of misdetection $(P_{MD} = 1 - P_D)$ for $U=0$ dB and $SNR=10$ dB. We can observe that P_{MD} increases as noise correlation ε increases from 0 to 1. As $\varepsilon \to 1, P_{MD} \to 1$ or $P_D \to 0$ and system shows a zero performance. In this figure, P_{MD} is plotted for different values of ε at $U=0.5$ dB and $d=1.035$.

Figure 5 is drawn at $P_{FA,des} = 0.1$ and $SNR=-10$ dB. We can see that lower ε is required for higher U for robust detection. Also, the introduced d factor improves performance of P_{MD}.

Figure 6 indicates the comparative analysis between the proposed method and the existing ones [3, 26, 28] in term of energy detection under the number of noise uncertainty. It is clearly indicated that the proposed work performs better by the rate of 20.08% when the values of N increases and $\varepsilon = 0.001$.

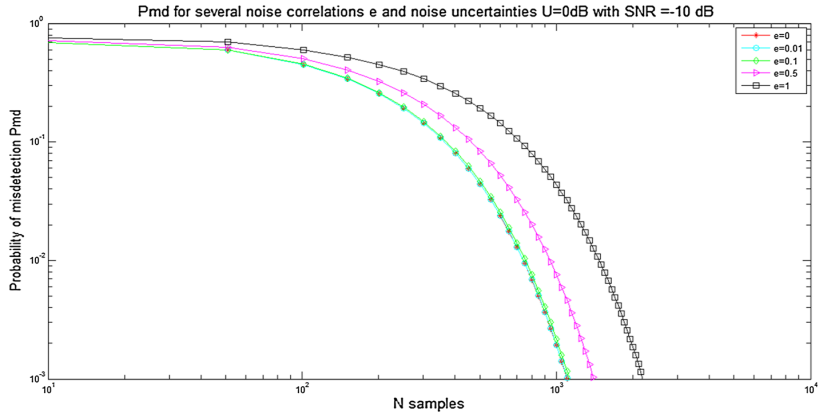


Fig. 4 Probability of misdetection (P_{MD}) by samples (N) by values of ε

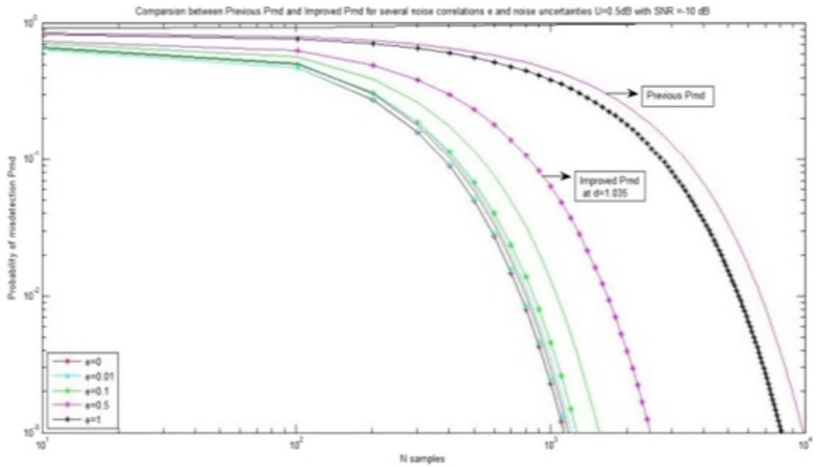


Fig. 5 Improved P_{MD} (at $d=1.035$) and the previous P_{MD} with $U=0.5$ dB, $SNR=-10$ dB [26, 28]. Improved P_{MD} at $d=1.035$. Previous P_{MD}

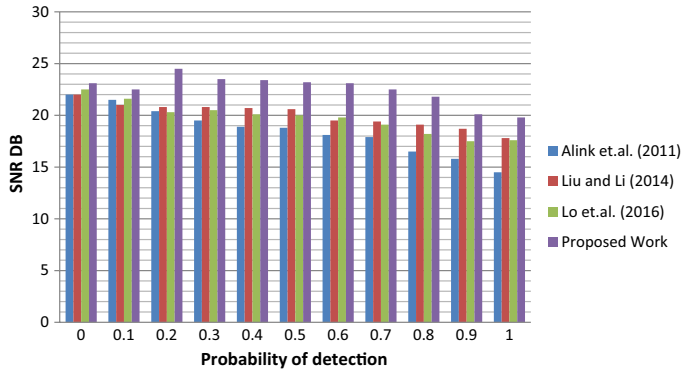


Fig. 6 Probability of energy detection under the number of noise uncertainty ($N=1000$, $\varepsilon=0.001$)

5 Conclusion

This article analyzed the performance of energy detector degraded when noise uncertainty is available. In this paper, we introduced a new hybrid method based on cross-correlation method and adaptive threshold method. An expression for SNR wall was derived which relates to correlation ε, uncertainty U and dynamic threshold factor d. In the cross-correlation, same signals pass to two different paths which reduce uncorrelated noises. Therefore, SNR is improved at outputs. Various simulations have been carried out to verify the accuracy of the derived relation on SNR degradation and packet loss charge. We identified various noises (Eqs. 13 to 24). Uncertainty yields out to derive a lower bound on the SNR Wall of the sensor networks were investigated. The results obtained from the previous techniques were compared with the proposed method. Simulations showed that new derived SNR wall improves detection performance compares to the correlation-based method. The experiments clearly affirmed the advantages of the proposed work.

Future studies will investigate an effective computational model that integrates different data sources in the proposed method [8, 13, 16, 38, 39]; Kapoor et al. [21, 22, 29, 32–35, 40–54].

References

1. Ali, K. A., Rehmani, H., & Reisslein, M. (2016). Cognitive radio for smart grids: Survey of architectures, spectrum sensing mechanisms, and networking protocols. *IEEE Communications Surveys & Tutorials, 18*(1), 860–898.
2. Ali, G., Qaraqe, K., Celebi H. &Arslan H. (2010) An adaptive threshold method for spectrum sensing in multi-channel cognitive radio networks. In *2010 IEEE 17th international conference on telecommunications (ICT 2010)*, (pp. 425–429), IEEE.
3. Alink, M. S. O., Kokkeler, A. B., Klumperink, E. A., Smit, G. J., & Nauta, B. (2011). Lowering the SNR wall for energy detection using cross-correlation. *IEEE Transactions on Vehicular Technology, 60*(8), 3748–3757.
4. Beaulieu, D. L. (2000). Comprehensive reform and American Indian education. *Journal of American Indian Education, 39*(2), 29–38.
5. Bruckner, D., Velik, R., & Penya, Y. (2011). Machine perception in automation: A call toarms. *EURASIP Journal on Embedded Systems, 6*(8), 1–9.
6. Claudino, L., & Abrão, T. (2017). Spectrum sensing methods for cognitive radio networks: A review. *Wireless Personal Communications, 95*(4), 5003–5037.
7. Dhurgadevi, M., & Devi, P. M. (2018). An analysis of energy efficiency improvement through wireless energy transfer in wireless sensor network. *Wireless Personal Communications, 98*(4), 3377–3391.
8. Doss, S., Nayyar, A., Suseendran, G., Tanwar, S., Khanna, A., Son, L. H., et al. (2018). APD-JFAD: Accurate prevention and detection of Jelly Fish attack in MANET. *Ieee Access, 6,* 56954–56965.
9. Dutta P. & Manna G.C. (2016) Designing a cognitive radio with enhancement in throughput and improved spectrum sensing technique. In *2nd IEEE international conference on control science and systems engineering (ICCSS 2016)*, (pp. 27–29) July 2016, Singapore. https://doi.org/10.1109/ccsse.2016.7784345.
10. Emara, M., et al. (2016). Spectrum sensing optimization and performance enhancement of cognitive radio networks. *Wireless Personal Communications, 86*(2), 925–941.
11. Gao, N., et al. (2017). Robust collaborative spectrum sensing using PHY-layer fingerprints in mobile cognitive radio networks. *IEEE Communications Letters*. https://doi.org/10.1109/lcomm.2017.2656901.
12. Gao, R., Li, Z., Li, H., & Ai, B. (2015). Absolute value cumulating based spectrum sensing with Laplacian noise in cognitive radio networks. *Wireless Personal Communications, 83*(2), 1387–1404.
13. Garg, R, Mittal, M, Son, LH (2019). Reliability and energy efficient workflow scheduling in cloud environment. Cluster Computing, (**in press**).
14. Gohain, P. B., Chaudhari, S., &Koivunen, V. (2018) Cooperative energy detection with heterogeneous sensors under noise uncertainty: SNR wall and use of evidence theory. IEEE Transactions on Cognitive Communications and Networking.
15. Gunes, N., Higgins, D. M., & Leeson, S. M. (2016). A stochastic resonator to detect BPAM signals; analysis, PSR designs, and sine-induced SR. *IET Signal Processing*. https://doi.org/10.1049/iet-spr.2015.0152.
16. Hai, D. T., Son, H., & Vinh, L. T. (2017). Novel fuzzy clustering scheme for 3D wireless sensor networks. *Applied Soft Computing, 54,* 141–149.
17. Haijun, Z., et al. (2016). Interference-limited resource optimization in cognitive femtocells with fairness and imperfect spectrum sensing. *IEEE Transactions on Vehicular Technology, 65*(3), 1761–1771.
18. Jha, S. K., & Eyong, E. M. (2018). An energy optimization in wireless sensor networks by using genetic algorithm. *Telecommunication Systems, 67*(1), 113–121.

19. Joshi, D., Dimitrie, P., & Octavia, D. (2011). Gradient-based threshold adaptation for energy detector in cognitive radio systems. *IEEE Communications Letters, 15*(1), 19–21.
20. Junaid, I., & Kim, D. (2017). Energy-Efficient Management of Cognitive Radio Terminals with Quality-Based Activation. *IEEE Communications Letters, 21*(5), 1171–1174.
21. Kapoor, R., Gupta, R., Jha, S., & Kumar, R. (2018). Detection of power quality event using histogram of oriented gradients and support vector machine. *Measurement, 120,* 52–75.
22. Kapoor, R., Gupta, R., Kumar, R., Son, L. H., & Jha, S. (2019). New scheme for underwater acoustically wireless transmission using direct sequence code division multiple access in MIMO systems. *Wireless Networks*. https://doi.org/10.1007/s11276-018-1750-z.
23. Kapoor, R., Gupta, R., Son, L. H., Jha, S., & Kumar, R. (2018). Detection of power quality event using histogram of oriented gradients and support vector machine. *Measurement, 120,* 52–75.
24. Kapoor, R., Gupta, R., Son, L. H., Jha, S., & Kumar, R. (2018). Boosting performance of power quality event identification with KL Divergence measure and standard deviation. *Measurement, 126,* 134–142.
25. Kapoor, R., Gupta, R., Son, LH, Kumar, R., & Jha, S. (2018b) New scheme for underwater acoustically wireless transmission using direct sequence code division multiple access in MIMO systems. Wireless Networks, pp. 1–13.
26. Liu, J., & Li, Z. (2014). Lowering the signal-to-noise ratio wall for energy detection using parameter-induced stochastic resonator. *IET Communications, 9*(1), 101–107.
27. Liu, J., Youguo, W., & Qiqing, Z. (2016). Stochastic resonance of signal detection in mono-threshold system using additive and multiplicative noises. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 99*(1), 323–329.
28. Lo, Y. S., Lim, H. S., & Tan, A. W. C. (2016). Robust signal-to-noise ratio estimation in non-gaussian noise channel. *Wireless Personal Communications, 91*(2), 561–575.
29. Long, H. V., Ali, M., Khan, M., & Tu, D. N. (2019). A novel approach for fuzzy clustering based on neutrosophic association matrix. *Computers & Industrial Engineering*. https://doi.org/10.1016/j.cie.2018.11.007.
30. Mashreghi, M., & Abolhassani, B. (2017). A cluster-based cooperative spectrum sensing strategy to maximize achievable throughput. *Wireless Personal Communications, 96*(3), 4557–4584.
31. Oude, A., et al. (2011). Lowering the SNR wall for energy detection using cross-correlation. *IEEE Transactions on Vehicular Technology, 60*(8), 3748–3757.
32. Phuong, P. T. M., Thong, P. H., & Son, L. H. (2018). Theoretical analysis of picture fuzzy clustering: Convergence and property. *Journal of Computer Science and Cybernetics, 34*(1), 17–32.
33. Robinson, Y. H., Julie, E. G., Saravanan, K., Kumar, R., & Son, L. H. (2019). FD-AOMDV: fault-tolerant disjoint ad-hoc on-demand multipath distance vector routing algorithm in mobile ad-hoc networks. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-018-1126-3.
34. Saravanan, K., Anusuya, E., Kumar, R., & Son, L. H. (2018). Real-time water quality monitoring using Internet of Things in SCADA. *Environmental Monitoring and Assessment, 190*(9), 556.
35. Saravanan, K., Aswini, S., Kumar, R., & Son, L. H. (2019). How to prevent maritime border collision for fisheries?-A design of Real-Time Automatic Identification System. *Earth Science Informatics, 1,* 1–12. https://doi.org/10.1007/s12145-018-0371-5.
36. Sarkar, S., Virani, N., Yasar, M., Ray A. & Sarkar S. (2013) Spatiotemporal information fusion for fault detection in shipboard auxiliary systems. American Control Conference, Washington D. C., (pp. 3846–3851).
37. Shaikh B., Zafi S., &Umrani F. (2016) An unsigned autocorrelation based blind spectrum sensing approach for cognitive radio. In *2016 IEEE International Conference on Open Source Systems & Technologies (ICOSST 2016)*, 15–17 Dec. 2016, Lahore, Pakistan. https://doi.org/10.1109/icosst.2016.7838576.
38. Singh, K., Singh, K., Son, L. H., & Aziz, A. (2018). Congestion control in wireless sensor networks by hybrid multi-objective optimization algorithm. *Computer Networks, 138,* 90–107.
39. Singh, N., Son, L. H., Chiclana, F., & Jean-Pierre, M. (2019). A new fusion of salp swarm with sine cosine for optimization of non-linear functions. *Engineering with Computers*. https://doi.org/10.1007/s00366-018-00696-8.
40. Son, L. H. (2015). A novel kernel fuzzy clustering algorithm for geo-demographic analysis. Information Sciences—Informatics and Computer Science. *Intelligent Systems, Applications: An International Journal, 317,* 202–223.
41. Son, L. H. (2016). Generalized picture distance measure and applications to picture fuzzy clustering. *Applied Soft Computing, 46,* 284–295.
42. Son, L. H., & Hai, P. V. (2016). A novel multiple fuzzy clustering method based on internal clustering validation measures with gradient descent. *International Journal of Fuzzy Systems, 18*(5), 894–903.

43. Son, L. H., Jha, S., Kumar, R., Chatterjee, J. M., & Khari, M. (2019). Collaborative handshaking approaches between internet of computing and internet of things towards a smart world: a review from 2009 to 2017. *Telecommunication Systems*. https://doi.org/10.1007/s11235-018-0481-x.
44. Son, L. H., & Tien, N. D. (2017). Tune up fuzzy C-means for big data: some novel hybrid clustering algorithms based on initial selection and incremental clustering. *International Journal of Fuzzy Systems, 19*(5), 1585–1602.
45. Son, L. H., & Tuan, T. M. (2016). A cooperative semi-supervised fuzzy clustering framework for dental X-ray image segmentation. *Expert Systems with Applications, 46,* 380–393.
46. Son, L. H., & Thong, P. H. (2017). Some novel hybrid forecast methods based on picture fuzzy clustering for weather nowcasting from satellite image sequences. *Applied Intelligence, 46*(1), 1–15.
47. Son, L. H., & Tuan, T. M. (2017). Dental segmentation from X-ray images using semi-supervised fuzzy clustering with spatial constraints. *Engineering Applications of Artificial Intelligence, 59,* 186–195.
48. Tam, N. T., Hai, D. T., Son, L. H., & Vinh, L. T. (2018). Improving lifetime and network connections of 3D wireless sensor networks based on fuzzy clustering and particle swarm optimization. *Wireless Networks, 24*(5), 1477–1490.
49. Tandra, R., & Sahai, A. (2005). Fundamental limits on detection in low SNR under noise uncertainty. In *Wireless Networks, Communications and Mobile Computing, 2005 International Conference on* (Vol. 1, pp. 464–469). IEEE.
50. Thanh, N. D., Ali, M., & Son, L. H. (2017). A novel clustering algorithm in a neutrosophic recommender system for medical diagnosis. *Cognitive Computation, 9*(4), 526–544.
51. Thong, P. H., & Son, L. H. (2016). Picture fuzzy clustering: a new computational intelligence method. *Soft Computing, 20*(9), 3549–3562.
52. Thong, P. H., & Son, L. H. (2016). A novel automatic picture fuzzy clustering method based on particle swarm optimization and picture composite cardinality. *Knowledge-Based Systems, 109,* 48–60.
53. Thong, P. H., & Son, L. H. (2016). Picture fuzzy clustering for complex data. *Engineering Applications of Artificial Intelligence, 56,* 121–130.
54. Tuan, T. M., Ngan, T. T., & Son, L. H. (2016). A novel semi-supervised fuzzy clustering method based on interactive fuzzy satisficing for dental X-ray image segmentation. *Applied Intelligence, 45*(2), 402–428.
55. Zulfikar, A. et al. (2017) Enhanced spectrum sensing based on Energy detection in cognitive radio network using adaptive threshold. In *2017 IEEE international conference on networking, systems and security (NSysS)*, Doi: https://doi.org/10.1109/nsyss.2017.7885815.

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



Rajiv Kapoor received M.E. and Ph.D. degree in Electronics & Communication Engineering from Delhi College of Engineering, Delhi University. Dr. Kapoor presently working as Professor in Electronics & Communication Engineering Department, AIACT&R (Govt. of NCT of Delhi). He has authored over 90 research papers in various renowned international journal and conferences. His primary research interests are machine learning, computer vision, signal and image processing.



Rashmi Gupta received M.E. and Ph.D. degree in Electronics & Communication Engineering from Delhi College of Engineering, Delhi University. Dr. Gupta held the position of Sr. Engineer in Calcom group of companies from 1991 to 1999, position of Lecturer in Electronics & Communication Engineering Department at Hindu Institute of Technology and Senior Lecturer in Maharaja Agrasen Intitute of Technology, Delhi from 1999 to 2003. Dr. Gupta presently working as Associate Professor in Electronics & Communication Engineering Department, AIACT&R (Govt. of NCT of Delhi). She has authored over 40 research papers in various renowned international journal and conferences. Her primary research interests are machine learning, computer vision, signal and image processing.



Le Hoang Son obtained the PhD degree on Mathematics—Informatics at VNU University of Science, Vietnam National University (VNU). He has been promoted to Associate Professor in Information Technology since 2017. Currently, Dr. Son works as a Professor at Vietnam National University, Hanoi and a Visting Researcher at Ton Duc Thang University, Ho Chi Minh city, Vietnam. His major field includes Artificial Intelligence, Data Mining, Soft Computing, Fuzzy Computing, Fuzzy Recommender Systems, and Geographic Information System. He is a member of International Association of Computer Science and Information Technology (IACSIT), Center for Applied Research in e-Health (eCARE), Vietnam Society for Applications of Mathematics (Vietsam). Dr. Son serves as Editorial Board of International Journal of Ambient Computing and Intelligence (IJACI, SCOPUS), Editorial Board of Vietnam Journal of Computer Science and Cybernetics (JCC), Associate Editor of International Journal of Engineering and Technology (IJET), Associate Editor of Neutrosophic Sets and Systems (NSS), and Associate Editor of Vietnam Research and Development on Information and Communication Technology (RD-ICT).



Sudan Jha was born on October 22 at the Capital city Kathmandu of Himalayan Country Nepal. He received proficiency in certificate level from Saint Xavier's College Kathmandu and then Bachelor of Engineering in Electronics Engineering from Motilal Nehru Regional College, Allahabad. Uttar Pradesh, India, in 2001; joined as a Lecturer in one of the premium and largest Engineering College and the first one in the private domain in Nepal ("Nepal Engineering College (NEC)"), where he got full sponsorship from the employer (NEC) to pursue "Masters in Computer Science". In due course of time, he chaired and organized 5 International Conferences, some of the proceedings of those conferences been published by Springer Verlag, World Science Series, Imperial Press London, etc. He was promoted to Assistant Professor in the Department of Computer Science and Engineering after completion of his Masters Degree and no sooner than later, he became the Head of the Computer Science and Engineering Department.



Raghvendra Kumar is working as Assistant Professor in Computer Science and Engineering Department at L.N.C.T Group of College Jabalpur, M.P. India. He received B. Tech. in Computer Science and Engineering from SRM University Chennai (Tamil Nadu), India, M. Tech. in Computer Science and Engineering from KIIT University, Bhubaneswar, (Odisha) India and Ph.D. in Computer Science and Engineering from Jodhpur National University, Jodhpur (Rajasthan), India. He has published 86 research papers in international/National journal and conferences including IEEE, Springer and ACM as well as serve as session chair, Co-chair, Technical program Committee members in many international and national conferences and serve as guest editors in many special issues from reputed journals (Indexed By: Scopus, ESCI). He also received best paper award in IEEE Conference 2013 and Young Achiever Award-2016 by IEAE Association for his research work in the field of distributed database. His researches areas are Computer Networks, Data Mining, cloud computing and Secure Multiparty Computations, Theory of Computer Science and Design of Algorithms. He authored 12 computer science books in field of Data Mining, Robotics, Graph Theory, and Turing Machine by IGI Global Publication, USA, IOS Press Netherland, Lambert Publication, Scholar Press, Kataria Publication, Narosa, Edupedia Publication, S. Chand Publication and Laxmi Publication.

Breast Cancer
https://doi.org/10.1007/s12282-018-00941-4

ORIGINAL ARTICLE

Breast cancer incidence, mortality and mortality-to-incidence ratio (MIR) are associated with human development, 1990–2016: evidence from Global Burden of Disease Study 2016

Rajesh Sharma[1]

Received: 7 August 2018 / Accepted: 17 December 2018
© The Japanese Breast Cancer Society 2019

Abstract

Objective To examine breast cancer burden in females using incidence, mortality and mortality-to-incidence ratio (MIR) and its association with human development.

Methods We employ the data of breast cancer in females from the Global Burden of Disease 2016 study for the period 1990 to 2016 for 102 countries. Human development is measured using the human development index (HDI). 5-year survival rate of breast cancer is proxied using the mortality-to-incidence ratio (MIR).

Findings Globally, breast cancer has claimed 535341 female lives and 1.7 million incident cases had surfaced in 2016. High incidence rates were observed in very high HDI countries led by the Netherlands (117.2/100,000), whereas the mortality rate was high in low/medium HDI countries led by Afghanistan (35.4/100,000). Breast cancer incidence has more than doubled in 60/102 countries, whereas deaths have doubled in 43/102 countries. Globally, breast cancer MIR decreased from 0.41 to 0.32 over 1990–2016 and displayed negative gradient with HDI ($r = -0.87$), indicating a low 5-year survival in less developed countries.

Conclusion Heterogeneity in breast cancer burden, as per human development, and increasing breast cancer incidence and low survival rates, indicated by MIR, call for broader human development, improving breast cancer awareness, and cost-effective screening and treatment in less developed countries.

Keywords Breast cancer · Incidence · Mortality · Mortality-to-incidence ratio (MIR)

Introduction

Breast cancer is a leading cancer burden in females, claiming 535341 lives and resulting in 14.8 million disability-adjusted life years (DALYs), and 1.7 million new cases of breast cancer arose in 2016 making it one of the biggest causes of mortality and morbidity in females, worldwide [1, 2]. The incidence of breast cancer is increasing not only in developed countries, but also in developing countries [2]. The survival rate of breast cancer, although improved, showed distinct and heterogenous patterns in different countries due to factors such as lack of screening, early-stage detection and cost-effectiveness of treatment [3].

Measuring and tracking survival rate of different cancers (including breast cancer) is costly and mortality-to-incidence ratio (MIR) is demonstrated to be quite a useful indicator of 5-year survival [4]. 5-year survival proxied by MIR has been examined for different neoplasia: prostate cancer [5], bladder cancer [6], gastric cancer [7], liver cancer [8] and colorectal cancer [9]. Previous studies examining 5-year survival rates using mortality-to-incidence ratio (MIR) examined the association between MIR and human development index or other development indicators using data of select geographical locations (mostly advanced countries) or employed shorter time frame.

In this study, we examine the past trends of incidence, mortality and mortality-to-incidence ratio of breast cancer in females employing data from the Global Burden of Disease 2016 study for the period 1990–2016 for 102 countries. Progress in breast cancer and other neoplasms is often associated with country-level per capita income and countries are often categorised as per World Bank classification into low-, lower-middle, upper-middle and high-income countries [10]. We, however, conjecture that a country's human development index—a composite measure of education (composite of mean years of schooling and expected years of schooling), income (gross national income (GNI) per capita) and health (life expectancy at birth)—is a better measure of its development and more broadly reflects the efficiency and efficacy of its health-care system. Figure 1 illustrates this point showing that MIR is predicted better by HDI (adj. $R^2 = 0.78$) than its individual components, such as life expectancy (adj. $R^2 = 0.65$), mean years of schooling (adj. $R^2 = 0.64$) and expected years of schooling (adj. $R^2 = 0.60$), and least by per capita income (adj. $R^2 = 0.48$).

Therefore, we examine the past trends of breast cancer burden in terms of incidence, mortality (both all-age and age-standardised) and MIR across countries as per their HDI, which is expected to provide a more accurate assessment of the association between breast cancer burden in a country and its development status.

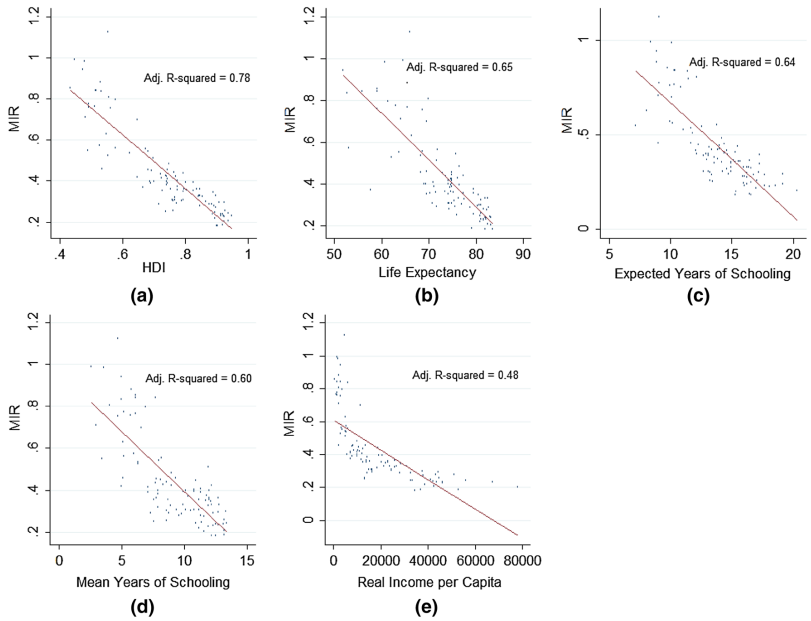


Fig. 1 Relationship between MIR and HDI components in 2015, **a** MIR vs HDI, **b** MIR vs LE, **c** MIR vs Exp schooling, **d** MIR vs mean schooling, **e** MIR vs income. *MIR* mortality-to-incidence ratio, *HDI* Human Development Index, *LE* life expectancy at birth (in years), *Exp schooling* expected years of schooling, *mean schooling* mean years of schooling, *income* real gross national income (GNI) per capita at 2011 purchasing power parity (PPP) prices. Data source: MIR was calculated by the author using crude mortality and incidence data of female breast cancer from the GBD 2016 study, and HDI data (and its components) was procured from the UNDP database

Materials and methods

We employ breast cancer data from the Global Burden of Disease (GBD) 2016 study, which provides location-, year-, age- and sex-wise estimates of cause of death for 195 countries and sub-national locations from 1990 to 2016 [1]. GBD 2016 employs all available data from sources such as vital registration, verbal autopsies, surveillance data and disease registry data, and uses Cause of Death Ensemble Model (CodeM) developed by Foreman et al. [11] to arrive at cause-specific mortality estimates. It uses separately estimated mortality-to-incidence ratio along with mortality estimates to arrive at cancer incidence [1, 2]. In this paper, we divide publicly available data of all-age crude death rate by all-age crude incidence rate of breast cancer in females from the GBD 2016 study to calculate the mortality-to-incidence

✉ Rajesh Sharma
rajesh.sharma@dtu.ac.in

[1] University School of Management and Entrepreneurship, Delhi Technological University, Vivek Vihar-Phase II, Delhi 110095, India

Published online: 02 January 2019

ratio (MIR). We employ the data of breast cancer in females for 102 countries, which had incidence greater than 1000 cases in 2016 [12].[1] Combined, these countries accounted for 97% of global cancer deaths and 98% of incident cases in 2016. GBD estimates are provided with 95% uncertainty intervals (UI) and are reported inside square brackets along with mean estimates in this paper; where percentage change is reported alone, it is calculated from the mean estimates.

The development level of a country is measured using Human Development Index (HDI): a composite measure of education (composite of mean years of schooling and expected years of schooling), income (GNI per capita) and health (life expectancy at birth) [13]. Each of these individual indicators is converted to a scale of 0 (worst) to 1 (best) and the geometric mean of the three individual indicators gives the aggregate value of HDI for a given country in a year. Data pertaining to Human Development Index (HDI) for the period 1990–2015 is procured from the United Nations Development Program (UNDP) database for 100 countries, except for North Korea and Puerto Rico which did not have HDI values [13]. For descriptive as well as statistical analysis, countries were categorised into four groups as per UNDP classification of countries in 2015: very high (HDI > 0.800, 36 countries), high (0.700 < HDI < 0.799, 31 countries), medium (0.550 < HDI < 0.669, 18 countries) and low (HDI < 0.550, 15 countries) HDI categories.[2] All the analysis is done in MS-Excel 2016 and Stata 13.

Results

Globally, female deaths from breast cancer increased from 336857 [95% UI 319076–367033] in 1990 to 535341 [95% UI 506372–573133] in 2016 (Table 1). The incident cases increased from 815406 [95% UI 788481–857562] in 1990 to 1.68 million [95% UI 1.61–1.78] in 2016. In age-standardised terms, globally, the rate of breast cancer mortality (age-standardised mortality rate, ASMR) decreased from 17.2 [95% UI 16.4–18.9] in 1990 to 14.6 [95% UI 13.8–15.6] per 100,000 females in 2016, whereas age-standardised incidence rate (ASIR) increased from 41.0 [39.8–43.1] in 1990 to 45.6 [95% UI 43.6–48.2] in 2016. 102 countries combined in this study accounted for 1.65 million of the total 1.68 million incident cases and 517614 of the total 535341 global incident cases of female breast cancer in 2016.

[1] 1000 cases are chosen so as to exclude countries with too few cancer cases as it may lead to too large or too small MIR values which may not truly reflect countries' development status and may distort main conclusions of the paper.

[2] Country-specific HDI values and component-wise values in 2015 are presented in Table 3 of the "Appendix".

HDI category-wise burden in 2016

In terms of absolute numbers, very high HDI locations led all HDI categories followed by high HDI locations with low and medium HDI countries recording lesser numbers (both incidence and mortality) as compared to their developed counterparts (col I-II, Fig. 2). In terms of incidence and mortality, China (high HDI country) was the leading country with 290726 [95% UI 232995–333019] incident cases and 69872 [95% UI 58945–79505] deaths due to breast cancer in 2016 (Table 1). Interestingly, India a medium HDI country, ranks third in terms of breast cancer incidence (115922 [95% UI 104676–127235]) behind China and the USA (very high HDI country, 259634 [95% UI 249312–270812]). This is in contrast to mortality, which is more reminiscent of the strength of a country's health-care system and its development, being higher in India (60462 [95% UI 51651–68759]) than in the USA (48780 [95% UI 46819–50963]), although the number of incident cases was higher in the USA than in India.

In terms of age-standardised incidence rate (ASIR), very high HDI locations occupied 19 of the top 20 positions, whereas in terms of age-standardised mortality rate (ASMR) 8 out of top 20 countries belonged to the very high HDI category (Table 1). The Netherlands and Belgium were the leading countries with ASIR of 117.2 [107.6–128.3] and 115.2 [95% UI 106.9–126.1] per 100,000 females, respectively, in 2016 (Fig. 2; Table 1). The age-standardised mortality rate (ASMR), however, was higher in the low HDI category than in other HDI categories with the highest ASMR recorded by Afghanistan at 35.4 [95% UI 14.6–64.5] per 100,000 females, followed by 31.5 [95% UI 19.0–51.3] per 100,000 females in Haiti in 2016. The age-standardised incidence rate (ASIR) was the lowest in North Korea at 13.1 [95% UI 9.6–19.7] with ASMR being the lowest in South Korea at 7.7 [4.8–11.4] per 100,000 females in 2016 (Table 1).

Mortality-to-incidence ratio (MIR), a representative indicator of 5-year survival, exhibited distinct patterns compared to absolute numbers and age-standardised rates with the highest MIR recorded in low HDI countries followed by medium HDI countries and the least MIR recorded in the very high HDI category (col v, Fig. 2). The majority of *very high* HDI locations recorded MIR lower than the global MIR of 0.32 in 2016, whereas the majority of low HDI countries were present at the bottom: 14 of 15 low HDI countries and 12 of the 18 medium HDI countries registered MIR greater than 0.5, whereas all countries in the high and very high HDI categories recorded MIR lower than 0.5 in 2016. In our sample of countries, Japan and South Korea recorded minimum MIR close to 0.18, closely followed by the USA with MIR of 0.19 in 2016. Country-wise, the highest MIR in the very high HDI category was recorded in Argentina with an MIR of 0.39, even greater than South Africa, a medium HDI country, which registered an MIR of 0.36. Among the high HDI locations, China

Table 1 Female breast cancer burden in 2016

Country	ASIR	ASMR	All age death numbers	All age incidence number	All age incidence rate	All age death rate	Mortality-to-incidence ratio
Low HDI							
Afghanistan	34.38	35.36	2689	2735	16.91	16.62	0.98
Angola	22.02	19.78	1032	1246	9.56	7.92	0.83
Cameroon	21.43	19.99	1077	1288	10.74	8.98	0.84
Cote d'Ivoire	24.87	25.57	1343	1433	12.64	11.84	0.94
Democratic Republic of the Congo	17.37	16.18	2788	3270	8.20	6.99	0.85
Ethiopia	15.60	16.71	4049	4101	8.00	7.90	0.99
Haiti	38.85	31.52	1101	1428	25.39	19.58	0.77
Madagascar	18.52	15.21	922	1217	9.73	7.37	0.76
Nigeria	38.15	25.23	11008	19132	21.10	12.14	0.58
Sudan	29.58	17.49	1916	3537	18.04	9.77	0.54
Syria	17.12	8.20	465	1046	11.62	5.16	0.44
Tanzania	13.80	13.40	1565	1797	6.56	5.71	0.87
Uganda	32.96	27.71	2038	2720	13.46	10.09	0.75
Yemen	26.63	20.00	1290	1855	13.32	9.27	0.70
Zimbabwe	28.28	26.72	870	1026	12.68	10.74	0.85
Medium HDI							
Bangladesh	14.70	8.40	4766	8824	11.00	5.94	0.54
Bolivia	26.81	14.97	600	1094	19.81	10.86	0.55
Cambodia	25.86	20.48	1184	1587	19.45	14.51	0.75
Egypt	23.65	9.49	3231	8427	18.57	7.12	0.38
Ghana	16.43	14.10	1091	1390	9.80	7.69	0.79
India	21.58	12.06	60462	115922	18.28	9.54	0.52
Indonesia	34.58	14.62	16294	41952	32.75	12.72	0.39
Iraq	35.89	26.33	2623	3785	19.48	13.50	0.69
Kenya	14.62	8.60	888	1706	7.32	3.81	0.52
Moldova	36.25	17.99	512	1017	48.11	24.21	0.50
Morocco	44.58	18.92	3094	7558	44.37	18.17	0.41
Myanmar	16.74	19.68	4692	4283	15.37	16.84	1.10
Nepal	14.43	12.30	1250	1552	10.03	8.08	0.81
Pakistan	42.51	28.96	16132	26103	28.11	17.37	0.62
Paraguay	46.93	21.12	506	1160	35.23	15.37	0.44
Philippines	45.99	22.30	8112	18145	35.72	15.97	0.45
South Africa	46.40	17.69	3835	10600	39.45	14.27	0.36
Vietnam	19.68	10.64	5195	9718	20.40	10.91	0.53
High HDI							
Algeria	31.13	14.43	2392	5460	27.27	11.95	0.44
Armenia	60.69	22.60	488	1261	76.89	29.76	0.39
Azerbaijan	37.49	11.84	622	2046	41.34	12.56	0.30
Belarus	42.16	14.14	1165	3212	62.87	22.80	0.36
Bosnia and Herzegovina	42.55	18.76	580	1238	64.74	30.35	0.47
Brazil	42.57	15.54	16716	46939	44.01	15.67	0.36
Bulgaria	67.39	18.70	1287	4204	113.44	34.74	0.31
China	32.54	8.15	67850	280951	42.35	10.23	0.24
Colombia	35.47	12.77	2955	8551	34.71	11.99	0.35
Costa Rica	42.40	14.84	375	1081	44.92	15.57	0.35
Cuba	50.74	18.03	1556	42vvvv20	74.31	27.40	0.37
Dominican Republic	28.66	12.65	550	1284	24.28	10.41	0.43
Ecuador	21.81	9.40	644	1539	18.58	7.77	0.42
Georgia	49.18	18.61	559	1377	65.39	26.55	0.41

Table 1 (continued)

Country	ASIR	ASMR	All age death numbers	All age incidence number	All age incidence rate	All age death rate	Mortality-to-incidence ratio
Iran	31.30	9.75	3083	11042	27.35	7.64	0.28
Jordan	49.88	21.38	474	1216	32.28	12.58	0.39
Kazakhstan	40.30	14.69	1352	3800	41.14	14.64	0.36
Lebanon	94.53	24.36	622	2451	84.44	21.42	0.25
Libya	41.53	13.33	320	1127	36.70	10.42	0.28
Malaysia	52.21	17.28	2118	7139	45.96	13.64	0.30
Mexico	40.49	12.35	6782	22979	35.54	10.49	0.30
Peru	22.22	9.69	1324	3140	19.47	8.21	0.42
Serbia	69.22	25.24	1865	4661	103.71	41.50	0.40
Sri Lanka	21.33	9.45	1104	2545	23.67	10.27	0.43
Thailand	31.12	11.76	5297	14145	41.29	15.46	0.37
Tunisia	41.75	17.50	1014	2504	44.13	17.87	0.40
Turkey	32.49	10.30	4149	13340	33.06	10.28	0.31
Ukraine	46.89	18.01	7283	17615	71.77	29.67	0.41
Uruguay	70.07	26.47	719	1662	93.41	40.38	0.43
Uzbekistan	32.72	13.26	1677	4382	28.35	10.85	0.38
Venezuela	52.72	16.47	2215	7382	46.73	14.02	0.30
Very high HDI							
Argentina	65.66	23.80	6403	16427	73.63	28.70	0.39
Australia	94.75	17.62	3202	15665	129.46	26.46	0.20
Austria	79.40	18.81	1613	5795	130.60	36.34	0.28
Belgium	115.21	24.15	2562	10392	180.14	44.41	0.25
Canada	93.00	19.82	6037	25472	139.05	32.96	0.24
Chile	40.46	12.59	1496	4584	49.82	16.26	0.33
Croatia	76.11	22.85	981	2777	127.11	44.91	0.35
Czech Republic	64.11	16.65	1652	5642	104.32	30.54	0.29
Denmark	98.49	25.21	1319	4471	155.30	45.80	0.29
Finland	90.99	16.32	882	4151	148.52	31.55	0.21
France	91.88	20.92	13041	47871	143.12	38.99	0.27
Germany	97.05	21.04	18430	71955	172.00	44.05	0.26
Greece	77.25	19.65	2311	7640	137.17	41.49	0.30
Hungary	69.86	22.39	2149	5963	115.08	41.48	0.36
Ireland	101.36	22.10	708	3072	130.75	30.12	0.23
Israel	75.27	22.97	1146	3463	83.99	27.79	0.33
Italy	100.46	19.49	13129	54568	174.83	42.06	0.24
Japan	59.09	9.35	11533	63065	97.71	17.87	0.18
Lithuania	55.57	17.45	504	1400	89.48	32.25	0.36
The Netherlands	117.16	23.37	3613	15900	183.81	41.77	0.23
New Zealand	98.57	22.04	749	3139	135.06	32.22	0.24
Norway	83.78	16.86	727	3137	120.36	27.89	0.23
Poland	60.35	17.26	5788	18144	90.94	29.01	0.32
Portugal	62.07	17.06	1849	5710	103.62	33.56	0.32
Romania	50.30	17.69	3054	7897	79.06	30.57	0.39
Russia	55.55	18.83	23534	64797	82.64	30.01	0.36
Saudi Arabia	36.43	12.01	878	3351	24.50	6.42	0.26
Singapore	75.24	14.49	390	1992	99.48	19.46	0.20
Slovakia	63.43	19.48	866	2635	94.25	30.98	0.33
Slovenia	65.71	18.52	391	1125	108.01	37.54	0.35
South Korea	44.43	7.67	2892	15995	63.15	11.42	0.18
Spain	59.87	14.88	6793	23466	99.13	28.70	0.29
Sweden	96.70	18.58	1717	7495	151.71	34.76	0.23

Table 1 (continued)

Country	ASIR	ASMR	All age death numbers	All age incidence number	All age incidence rate	All age death rate	Mortality-to-incidence ratio
Switzerland	88.22	18.86	1457	5941	140.48	34.45	0.25
United Kingdom	89.53	22.93	13367	45229	136.46	40.33	0.30
USA	111.62	19.45	48780	259634	158.43	29.77	0.19
No HDI category							
Global	45.57	14.61	535341	1681935	45.87	14.60	0.32
North Korea	13.12	9.29	1424	2016	14.94	10.55	0.71
Puerto Rico	70.03	17.01	472	1804	94.62	24.73	0.26

Countries were categorised into four groups as per HDI value in 2015: very high (HDI > 0.800), high (0.700 < HDI < 0.799), medium (0.550 < HDI < 0.669) and low (HDI < 0.550). Data Source: Global burden of disease 2016 study [12]

ASIR age-standardised incidence rate, *ASMR* age-standardised mortality rate

recorded the lowest MIR of 0.24 in 2016 (which is even lower than many very high HDI countries), and Bosnia and Herzegovina (high HDI country) recorded the highest MIR in its category (0.47)—even greater than the MIR of few medium HDI countries. Myanmar, a medium HDI country, was the only country with an MIR greater than 1 in 2016, which was even greater than the MIR of many low HDI countries.

Temporal patterns: 1990–2016

Breast cancer exhibited distinct patterns of change over the period 1990 to 2016 when examined using different metrics: all age incidence and mortality, age-standardised rates and mortality-to-incidence ratio (MIR). Breast cancer mortality more than doubled in 43 countries out of 102 countries (more than tripled in ten countries) with Iran recording the maximum increase of 289% between 1990 and 2016 (Table 2 of "Appendix"). In terms of incidence, breast cancer burden more than tripled in 30 countries with 30 other countries experiencing more than the doubling of breast cancer incidence over the period 1990–2016; the largest increase of 855% was registered in Saudi Arabia, whereas the minimum change was experienced by Georgia where it decreased by 18.5% between 1990 and 2016. The breast cancer burden in already heavily burdened countries of China, India and the USA increased by 84, 130 and 12%, respectively, in terms of incidence and increased by 286, 188 and 40%, respectively, in terms of mortality over the period 1990–2016.[3]

Looking at the temporal patters of ASMR, percent changes are not as big as that observed with respect to all-age incidence and mortality. 40 countries experienced a rise in ASMR between 1990 and 2016, and *very high* and high HDI locations underwent a reduction in ASMR barring few exceptions: ASMR increased by 32% in Saudi Arabia (*very high* HDI) and increased by 61% in Iran (high HDI location) (Table 2 in "Appendix"). ASMR decreased the most in Switzerland and the UK by 42 and 36%, respectively, over the period 1990 to 2016. ASIR decreased the most in Nepal by 11% from 1990 to 2016 and increased in 86 of 102 countries, with five witnessing a doubling of ASIR and Saudi Arabia and Iran witnessing a tripling of ASIR between 1990 and 2016 (Table 2 in "Appendix").

In 102 countries, the incidence of breast cancer increased faster than mortality which resulted in the reduction of MIR in all but six countries over the period 1990–2016. The maximum reduction in MIR was recorded in Saudi Arabia from 0.72 in 1990 to 0.26 in 2016, followed by Libya and China, which experienced a reduction close to 53% between 1990 and 2016. The six countries (five in low HDI category) which experienced an increase in MIR were: Zimbabwe (13.8%), North Korea (11%), Angola (4.6%), Democratic Republic of Congo (3.9%), Ethiopia (2.73%) and Afghanistan (2.43%).

Bivariate association between MIR and HDI

MIR of breast cancer showed a negative gradient with HDI—low HDI countries recorded the highest MIR both in 1990 and 2016, followed by medium HDI countries, while *very high* HDI countries registered the lowest MIR with few exceptions (Fig. 3) (pairwise correlation, $r = -0.87$). Over the period 1990–2015, as HDI increased for different countries, MIR of breast cancer decreased correspondingly, showing the significance of human development in reducing breast cancer burden and improving breast cancer survival. Pairwise correlation between MIR and HDI was the maximum in case of *very high* HDI countries with correlation coefficient of −0.78, followed by medium HDI countries ($r = -0.62$) and high HDI countries ($r = -0.58$) with least correlation recorded for low HDI countries ($r = -0.55$).

[3] Annual percentage change of incidence, mortality, ASIR, ASMR and MIR over the period 1990 to 2016 for different HDI groupings is shown in Fig. 5 of the "Appendix".

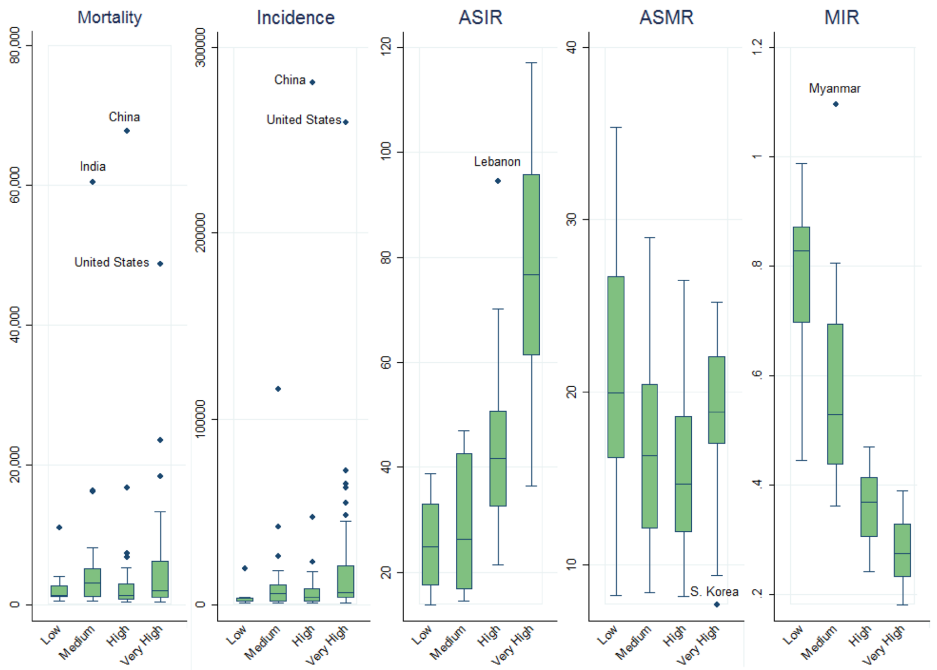


Fig. 2 HDI category-wise mortality, incidence, ASIR, ASMR and MIR in 2016. *ASIR* age-standardised incidence rate, *ASMR* age-standardised mortality rate, *MIR* mortality-to-incidence ratio. Data pertains to low HDI (15 countries), medium HDI (18 countries), high HDI (31 countries) and very high HDI (36 countries) for the period 1990 to 2016 and is procured from the Global Burden of Disease study 2016. Countries were categorised into four groups as per HDI value in 2015: very high (HDI > 0.800), high (0.700 < HDI < 0.799), medium (0.550 < HDI < 0.669) and low (HDI < 0.550)

Discussion

In this paper, we examine the breast cancer incidence and mortality in females for 102 countries for the period 1990–2016 and use mortality-to-incidence ratio (MIR) as a representative indicator of breast cancer survival. We investigate a country's progress in the fight against breast cancer vis-à-vis its development status, measured using the Human Development Index (HDI).

We observe that the global burden of breast cancer in females increased in terms of incidence and mortality between 1990 and 2016 with age-standardised rates either stagnated or decreased in the majority of countries examined in this study. The incidence (both all-age as well as age-standardised rate) of breast cancer is the highest in *very high* and *high* HDI countries, and correlates positively with country-specific HDI ($r = 0.77$). Breast cancer incidence more than tripled in 30 countries, whereas cancer mortality more than doubled in 43 out of 102 countries. Saudi Arabia recorded a maximum increase in incidence of 855% in the study period, whereas Iran registered a maximum increase of 289% in breast cancer mortality between 1990 and 2016 (Table 2 of "Appendix"). The incidence rates were the highest in developed countries led by the Netherlands and Belgium with ASIR of 117.2/100,000 and 115.2/100,000 females, respectively, in 2016. The age-standardised mortality rate (ASMR), however, was highest in the low HDI category led by Afghanistan with ASMR of 35.4/100,000 females, followed by 31.5/100,000 females in Haiti in 2016.

The breast cancer incidence, although highest in developed countries, is also on the rise in low and medium HDI countries, exemplified by countries such as India, a medium HDI country, which ranks third in terms of incidence (behind only China and the USA) and second behind China in terms of all-age mortality. What explains this rising incidence of

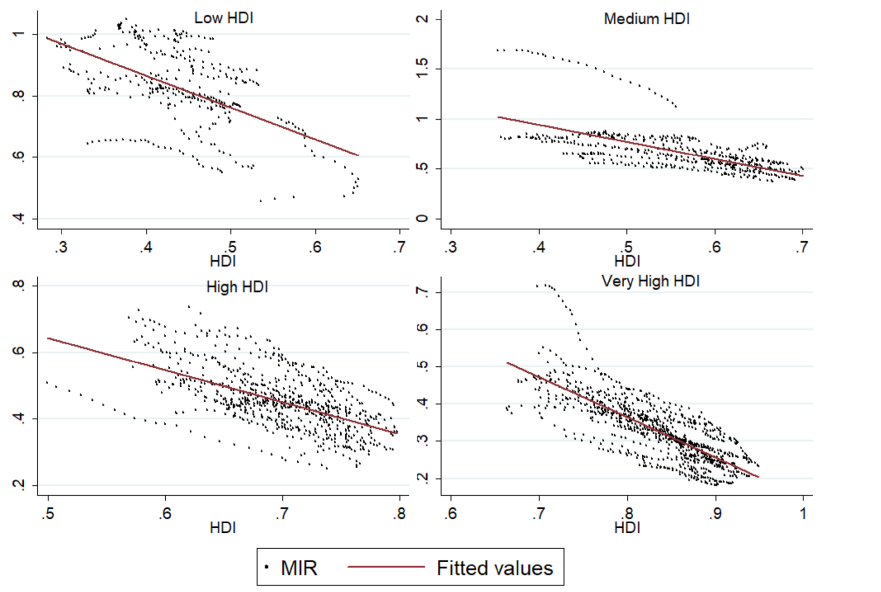


Fig. 3 MIR vs HDI over the period 1990–2015. Mortality-to-incidence ratio (MIR) is calculated as the ratio of crude death rate and crude incidence rate. Countries were categorised into four groups as per HDI value in 2015: very high (HDI > 0.800), high (0.700 < HDI < 0.799), medium (0.550 < HDI < 0.669) and low (HDI < 0.550)

breast cancer in low and medium HDI countries? There are multiple factors behind the increasing incidence of breast cancer in these countries: increased longevity, improved cancer registration and improving awareness towards breast cancer. Other likely reasons for increased incidence of breast cancer in developing countries is westernisation of lifestyle, characterised by a change in reproductive patterns such as earlier age at menarche, later first birth, lower parity, shorter duration of breastfeeding [14], and obesity and overweight [15]. The problem of the growing incidence of breast cancer in low and medium HDI countries is also associated with low survival rates [3] which is also evidenced by high MIR in these countries.

Globally, MIR reduced from 0.41 in 1990 to 0.32 in 2016; though all HDI categories underwent a reduction, high HDI locations experienced the steepest fall with low HDI countries recording the least progress in 5-year breast cancer survival measured using MIR (Fig. 6 in "Appendix"). Human development measured using HDI is strongly correlated with MIR implying that all the factors (health, education and income) incorporated in HDI, and not a single factor such as per capita income, are significant in determining the survival of breast cancer patients. On average, it is expected that health-care system's quality and access are similar in countries having similar HDI value and it also proved to be a good predictor of breast cancer survival in this study ($R^2 = 0.76$). In our analysis, however, there were a few exceptions to this relationship, with few countries performing better or worse than predicted by their HDI. For instance, South Korea in *very high* HDI category, China in high HDI category, Indonesia in medium HDI category and Nigeria in low HDI category performed better in terms of MIR than that predicted by their HDI, whereas Argentina and Russia in very high, Bosnia and Herzegovina in high, Myanmar in medium, and Ethiopia and Afghanistan in low HDI category performed worse than that predicted by their HDI. It implies that factors other than longevity, years of schooling and per capita income also explain the survival rate of breast cancer. Besides HDI, the survival rate of breast cancer depends upon country-specific healthcare factors such as access to screening, early detection, case-specific treatment (Her-2 positive or ER positive etc) and availability of infrastructure (radiological instruments, physicians etc). South Korea, for instance, having the lowest ASMR as well as low MIR (indicating high survival rate), started National Cancer Screening Programme in

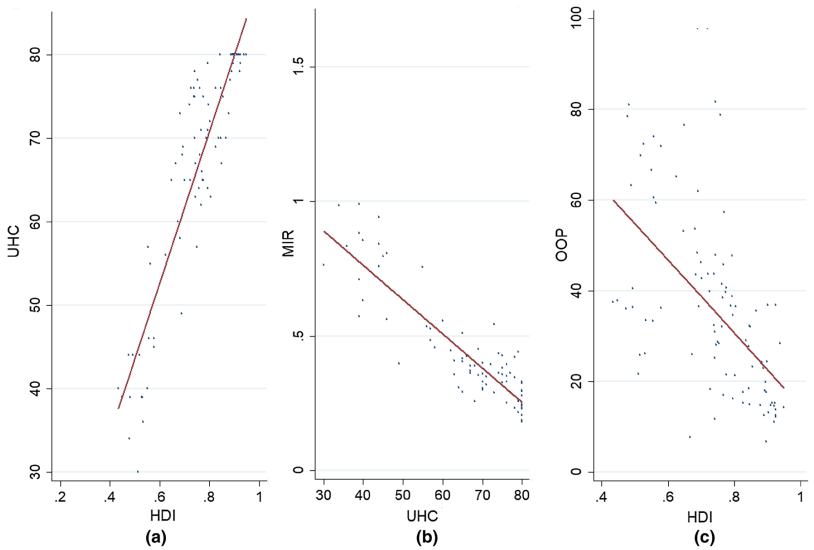


Fig. 4 Relationship between MIR and HDI with UHC: **a** UHC vs HDI, **b** MIR vs UHC, **c** OOP vs HDI. *MIR* mortality-to-incidence ratio, *HDI* Human Development Index, *UHC* universal health coverage service index, *OOP* out-of-pocket expenditure as percent of current health expenditure. Data source: MIR is calculated by the author using crude mortality and incidence data, HDI data is procured from the UNDP database and UHC data is procured from the WDI database of World Bank which in turn was compiled from Hogan et al. [30]. OOP data is also from the WDI database of World Bank

2002, which resulted in better prognosis of breast cancer and thereby higher survival rate [16].[4]

The stage of breast cancer at first diagnosis crucially determines its prognosis and differs markedly across countries—about 75% women in developing countries are diagnosed first in stage III and IV, whereas 70% of women in North America are diagnosed in stage 0 or 1, which results in better prognosis in developed regions as compared to the developing ones [3, 17, 18]. As far as early detection is concerned, WHO identified two strategies: *screening* and *clinical downstaging* [19]. Screening programs focus on detecting cancer in asymptomatic population, whereas clinical downstaging aims at ensuring that a patient with early symptoms meets an oncologist. In developed countries, breast cancer screening rates are high and screening has been quite instrumental in detecting cancer early and thereby improving survival in developed countries; the screening programs, however, are also found to be susceptible to high false positive rate and overdiagnosis/overtreatment [20–24]. Moreover, the cost-effectiveness of population-wide screening mammography is further questioned in developing countries; considering the low incidence rate per 100,000 females (due to large population) in these countries, the cost per life saved due to population-wide screening therefore, becomes astronomically high in these already resource-constrained economies [25]. Other alternatives to screening mammography are breast self-examination (BSE) and clinical breast examination (CBE). These methods are a cost-effective alternative to screening mammography; their efficacy, however, in early diagnosis and improved survival is still inconclusive [25, 26]. WHO's second early-detection strategy which calls for clinical downstaging to improve survival rates seems to be a more suitable and cost-effective alternative to screening mammography in low resource economies [27, 28].

Lastly, the high MIR and ASMR in low and medium HDI countries may also be reflective of other challenges specific to these countries. *First*, universal health coverage is positively correlated with HDI (Fig. 4a) and is negatively correlated with mortality-to-incidence ratio (MIR; Fig. 4b). Due to lack of universal health coverage in low and medium HDI countries, a large percentage of health expenditure incurred is spent from one's own pocket (Fig. 4c), which is also evidenced by high out-of-pocket (OOP) expenditure in low and medium HDI countries [29]. *Second*, most of the cancer care units in low and medium HDI countries are located in urban areas, making it more difficult and costlier for rural patients to access quality care. *Third*, there is low awareness (knowledge) of breast cancer. For instance, low percent of women in these countries are aware that a painless lump on the breast may be symptomatic of breast cancer, and if suspicious these women first reach out to a local healer after detecting any abnormality on the breast or nipple which causes a delay in diagnosis and thereby poor prognosis ensues. Therefore, improving breast cancer survival in these countries require strategies that focus on increased awareness, universal health coverage, improved access to either publicly funded or affordable screening programs and treatment paradigms.

[4] National Cancer Screening Program in Georgia also one of the successful screening program which resulted in downstaging of breast cancer and improved survival rate. Source: http://www.gnsc.ge/?act=page&id=44&lang=en (Accessed 18 Oct 2018).

Conclusion

This paper examines breast cancer burden in females in 102 countries for the period 1990–2016. We find that the incidence of breast cancer is higher in developed countries (high/very high HDI), but is also increasing in low and medium HDI countries. Breast cancer mortality rates were the highest in low and medium HDI countries, which was also supported by high MIR (proxy for survival rate) in these countries. Based on past trends, breast cancer incidence is expected to escalate further in future posing multifaceted challenges to already resource-constrained least developed countries. In the fight against breast cancer, a multipronged strategy must be adopted which focuses on improved human development, increases breast cancer awareness, universal health coverage and, lastly, breast cancer screening and treatment approaches which suit low-resource environments.

Limitations

This study employed data available from the Global Burden of Disease (GBD) study, which draws strength from data availability from cancer registries. The data from cancer registries, however, is far from complete in many resource-poor economies. This incomplete data yields mortality and incidence estimates with wide uncertainty intervals (UIs). This limitation, however, can be remedied by improved cancer registration in low/medium income countries which will help us gauge the current state of fight against cancer more accurately.

Compliance with ethical standards

Conflict of interest The authors of the paper declare no conflict of interest.

Appendix

See Figs. 5 and 6 and Tables 2 and 3.

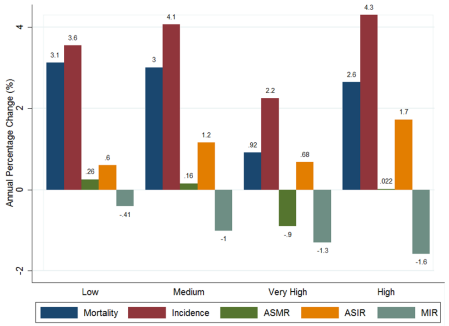


Fig. 5 HDI groupwise annual percentage change of breast cancer burden, 1990–2016. *ASIR* age-standardised incidence rate, *ASMR* age-standardised mortality rate, *MIR* mortality-to-incidence ratio. Countries were categorised into four groups as per HDI value in 2015: very high (HDI > 0.800), high (0.700 < HDI < 0.799), medium (0.550 < HDI < 0.669) and low (HDI < 0.550)

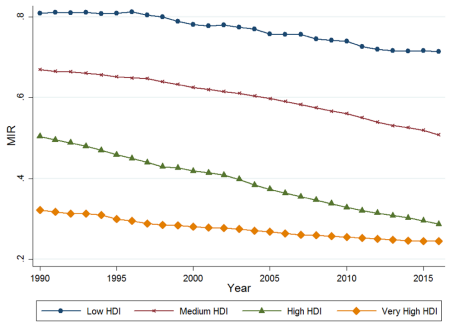


Fig. 6 HDI category-wise temporal movement of mortality-to-incidence ratio (MIR), 1990–2016. Data pertains to aggregate of data for low HDI (15) countries, medium HDI (18) countries, high HDI (31) countries and very high HDI (36) countries for the period 1990 to 2016 and is procured from Global Burden of Disease study 2016. Countries were categorised into four groups as per HDI value in 2015: very high (HDI > 0.800), high (0.700 < HDI < 0.799), medium (0.550 < HDI < 0.669) and low (HDI < 0.550)

Table 2 Total percentage change and annual percentage change (APC) of mortality, incidence, ASIR, ASMR and MIR of all countries over 1990–2016

Location name	Mortality		Incidence		ASMR		ASIR		MIR	
	Total change	APC	Total change	APC	Total change	APC	Total change	APC	Total change	APC
Low HDI										
Afghanistan	253.54	4.98	245.14	4.88	22.98	0.80	21.43	0.75	2.43	0.09
Angola	145.44	3.51	134.58	3.33	14.00	0.51	7.39	0.27	4.63	0.17
Cameroon	105.16	2.80	147.68	3.55	14.93	0.54	33.87	1.13	−17.17	−0.72
Cote d'Ivoire	135.36	3.35	158.24	3.72	19.11	0.67	28.96	0.98	−8.86	−0.36
Democratic Republic of the Congo	108.68	2.87	100.85	2.72	2.61	0.10	−2.31	−0.09	3.90	0.15
Ethiopia	95.01	2.60	89.82	2.50	−8.45	−0.34	−11.14	−0.45	2.74	0.10
Haiti	87.70	2.45	97.53	2.65	4.99	0.19	9.94	0.37	−4.98	−0.20
Madagascar	132.45	3.30	156.82	3.69	−1.43	−0.06	7.10	0.26	−9.49	−0.38
Nigeria	35.25	1.17	75.89	2.20	−23.61	−1.03	−5.64	−0.22	−23.10	−1.01
Sudan	145.18	3.51	191.61	4.20	4.30	0.16	22.50	0.78	−15.92	−0.66
Syria	149.06	3.57	307.55	5.55	10.58	0.39	77.56	2.23	−38.89	−1.88
Tanzania	126.61	3.20	168.12	3.87	6.85	0.26	24.30	0.84	−15.48	−0.64
Uganda	127.20	3.21	168.03	3.86	13.43	0.49	31.53	1.06	−15.23	−0.63
Yemen	210.97	4.46	268.43	5.14	25.72	0.88	45.04	1.44	−15.60	−0.65
Zimbabwe	67.43	2.00	47.17	1.50	8.55	0.32	−7.69	−0.31	13.76	0.50
Medium HDI										
Bangladesh	80.41	2.30	185.04	4.11	−17.91	−0.76	32.40	1.09	−36.71	−1.74
Bolivia	94.63	2.59	147.94	3.55	−8.95	−0.36	19.70	0.69	−21.50	−0.93
Cambodia	154.03	3.65	176.91	4.00	1.73	0.07	11.73	0.43	−8.26	−0.33
Egypt	103.94	2.78	241.03	4.83	8.62	0.32	78.23	2.25	−40.20	−1.96
Ghana	162.36	3.78	184.11	4.10	13.45	0.49	22.28	0.78	−7.66	−0.31
India	129.69	3.25	188.15	4.15	11.67	0.43	39.13	1.28	−20.29	−0.87
Indonesia	93.20	2.57	163.68	3.80	−3.08	−0.12	28.52	0.97	−26.73	−1.19
Iraq	137.81	3.39	178.66	4.02	−0.91	−0.04	13.22	0.48	−14.66	−0.61
Kenya	121.78	3.11	177.23	4.00	−5.82	−0.23	11.94	0.43	−20.00	−0.85
Moldova	11.09	0.41	18.92	0.67	−8.50	−0.34	−0.81	−0.03	−6.58	−0.26
Morocco	149.07	3.57	238.33	4.80	14.85	0.53	56.60	1.74	−26.38	−1.17
Myanmar	47.74	1.51	126.90	3.20	−20.99	−0.90	18.28	0.65	−34.89	−1.64
Nepal	68.84	2.04	71.24	2.09	−11.02	−0.45	−11.19	−0.46	−1.40	−0.05
Pakistan	194.22	4.24	266.74	5.12	35.95	1.19	65.54	1.96	−19.77	−0.84
Paraguay	236.99	4.78	399.28	6.38	52.79	1.64	130.04	3.26	−32.51	−1.50
Philippines	228.72	4.68	341.41	5.88	35.86	1.19	82.93	2.35	−25.53	−1.13
South Africa	125.12	3.17	208.88	4.43	21.80	0.76	67.62	2.01	−27.12	−1.21

Table 2 (continued)

Location name	Mortality		Incidence		ASMR		ASIR		MIR	
	Total change	APC	Total change	APC	Total change	APC	Total change	APC	Total change	APC
Vietnam	86.46	2.43	183.35	4.09	−14.04	−0.58	29.79	1.01	−34.19	−1.60
High HDI										
Algeria	241.13	4.83	464.70	6.88	31.89	1.07	117.68	3.04	−39.59	−1.92
Armenia	37.03	1.22	78.59	2.26	0.18	0.01	37.48	1.23	−23.27	−1.01
Azerbaijan	85.01	2.39	172.75	3.93	−6.17	−0.24	37.23	1.22	−32.17	−1.48
Belarus	1.36	0.05	21.32	0.75	−17.62	−0.74	1.74	0.07	−16.45	−0.69
Bosnia and Herzegovina	80.51	2.30	154.97	3.67	28.17	0.96	96.14	2.62	−29.21	−1.32
Brazil	126.88	3.20	237.94	4.79	−7.95	−0.32	44.03	1.41	−32.87	−1.52
Bulgaria	16.60	0.59	46.76	1.49	−1.24	−0.05	33.70	1.12	−20.55	−0.88
China	85.67	2.41	290.07	5.37	−10.42	−0.42	91.63	2.53	−52.40	−2.81
Colombia	100.05	2.70	195.27	4.25	−18.92	−0.80	22.07	0.77	−32.25	−1.49
Costa Rica	153.14	3.64	261.91	5.07	−7.28	−0.29	38.87	1.27	−30.06	−1.37
Cuba	73.86	2.15	110.09	2.90	−4.74	−0.19	18.75	0.66	−17.24	−0.73
Dominican Republic	176.32	3.99	290.90	5.38	14.97	0.54	70.48	2.07	−29.31	−1.33
Ecuador	137.42	3.38	229.67	4.70	−2.28	−0.09	39.23	1.28	−27.98	−1.25
Georgia	−23.82	−1.04	−18.52	−0.78	−17.42	−0.73	−7.77	−0.31	−6.51	−0.26
Iran	288.62	5.36	671.35	8.17	61.15	1.85	205.10	4.38	−49.62	−2.60
Jordan	219.95	4.57	503.67	7.16	4.38	0.17	84.44	2.38	−47.00	−2.41
Kazakhstan	28.98	0.98	72.32	2.11	2.75	0.10	34.98	1.16	−25.15	−1.11
Lebanon	122.66	3.13	362.48	6.07	−15.55	−0.65	80.04	2.29	−51.86	−2.77
Libya	203.36	4.36	548.25	7.45	20.44	0.72	145.10	3.51	−53.20	−2.88
Malaysia	151.69	3.61	424.95	6.59	−2.52	−0.10	99.42	2.69	−52.05	−2.79
Mexico	151.04	3.60	298.15	5.46	8.84	0.33	74.41	2.16	−36.95	−1.76
Peru	91.67	2.53	134.36	3.33	−16.39	−0.69	6.96	0.26	−18.22	−0.77
Serbia	36.67	1.21	54.70	1.69	2.34	0.09	27.01	0.92	−11.66	−0.48
Sri Lanka	133.24	3.31	266.14	5.12	12.74	0.46	87.16	2.44	−36.30	−1.72
Thailand	174.28	3.96	362.86	6.07	20.20	0.71	114.00	2.97	−40.74	−1.99
Tunisia	114.51	2.98	272.91	5.19	−4.92	−0.19	66.85	1.99	−42.48	−2.10
Turkey	79.01	2.26	266.47	5.12	−17.11	−0.72	70.08	2.06	−51.15	−2.72
Ukraine	−4.42	−0.17	−0.01	0.00	−12.46	−0.51	−6.42	−0.26	−4.40	−0.17
Uruguay	10.10	0.37	50.54	1.59	−23.44	−1.02	15.47	0.55	−26.87	−1.20
Uzbekistan	159.14	3.73	312.77	5.60	23.21	0.81	90.76	2.52	−37.22	−1.77
Venezuela	204.17	4.37	359.23	6.04	17.81	0.63	84.71	2.39	−33.77	−1.57

Table 2 (continued)

Location name	Mortality		Incidence		ASMR		ASIR		MIR	
	Total change	APC	Total change	APC	Total change	APC	Total change	APC	Total change	APC
Very high HDI										
Argentina	48.19	1.52	109.75	2.89	−10.97	−0.45	34.38	1.14	−29.35	−1.33
Australia	33.01	1.10	105.17	2.80	−30.34	−1.38	15.39	0.55	−35.17	−1.65
Austria	5.93	0.22	27.90	0.95	−25.01	−1.10	−4.88	−0.19	−17.18	−0.72
Belgium	−1.49	−0.06	32.84	1.10	−30.50	−1.39	0.96	0.04	−25.85	−1.14
Canada	28.41	0.97	70.01	2.06	−31.92	−1.47	−3.07	−0.12	−24.47	−1.07
Chile	79.36	2.27	195.31	4.25	−21.64	−0.93	39.05	1.28	−39.26	−1.90
Croatia	25.11	0.87	32.76	1.10	−4.67	−0.18	13.93	0.50	−5.76	−0.23
Czech Republic	0.63	0.02	25.25	0.87	−28.73	−1.29	−4.84	−0.19	−19.65	−0.84
Denmark	−10.46	−0.42	34.19	1.14	−34.35	−1.61	3.93	0.15	−33.27	−1.54
Finland	14.36	0.52	70.68	2.08	−24.63	−1.08	24.74	0.85	−33.00	−1.53
France	15.40	0.55	62.65	1.89	−25.17	−1.11	14.42	0.52	−29.05	−1.31
Germany	2.45	0.09	52.47	1.64	−25.00	−1.10	21.87	0.76	−32.81	−1.52
Greece	48.31	1.53	99.69	2.70	−11.39	−0.46	34.99	1.16	−25.73	−1.14
Hungary	2.54	0.10	33.25	1.11	−20.83	−0.89	10.91	0.40	−23.05	−1.00
Ireland	11.05	0.40	101.74	2.74	−34.45	−1.61	20.70	0.73	−44.96	−2.27
Israel	58.39	1.78	113.57	2.96	−36.11	−1.71	−7.51	−0.30	−25.84	−1.14
Italy	26.45	0.91	46.30	1.47	−19.49	−0.83	5.33	0.20	−13.57	−0.56
Japan	73.37	2.14	121.68	3.11	13.75	0.50	65.33	1.95	−21.79	−0.94
Lithuania	16.32	0.58	34.77	1.15	−3.44	−0.13	23.62	0.82	−13.69	−0.56
The Netherlands	11.60	0.42	63.60	1.91	−28.60	−1.29	12.18	0.44	−31.79	−1.46
New Zealand	34.14	1.14	78.04	2.24	−26.41	−1.17	0.70	0.03	−24.66	−1.08
Norway	−8.36	−0.34	58.68	1.79	−32.61	−1.51	22.01	0.77	−42.25	−2.09
Poland	26.68	0.91	87.97	2.46	−16.08	−0.67	35.40	1.17	−32.61	−1.51
Portugal	10.41	0.38	44.74	1.43	−31.02	−1.42	0.18	0.01	−23.72	−1.04
Romania	26.78	0.92	55.57	1.71	1.84	0.07	34.42	1.14	−18.51	−0.78
Russian Federation	45.46	1.45	59.44	1.81	14.00	0.51	28.18	0.96	−8.77	−0.35
Saudi Arabia	249.61	4.93	855.24	9.07	32.17	1.08	209.82	4.45	−63.40	−3.79
Singapore	113.55	2.96	271.28	5.17	−12.05	−0.49	64.27	1.93	−42.48	−2.10
Slovakia	32.95	1.10	95.43	2.61	−11.95	−0.49	36.77	1.21	−31.97	−1.47
Slovenia	27.45	0.94	38.55	1.26	−21.84	−0.94	0.46	0.02	−8.01	−0.32
South Korea	115.92	3.00	272.35	5.19	−1.90	−0.07	85.71	2.41	−42.01	−2.07

Table 2 (continued)

Location name	Mortality		Incidence		ASMR		ASIR		MIR	
	Total change	APC	Total change	APC	Total change	APC	Total change	APC	Total change	APC
Spain	18.06	0.64	55.69	1.72	−33.06	−1.53	−3.17	−0.12	−24.17	−1.06
Sweden	3.43	0.13	48.68	1.54	−24.04	−1.05	18.14	0.64	−30.44	−1.39
Switzerland	−12.70	−0.52	27.82	0.95	−42.11	−2.08	−10.93	−0.44	−31.71	−1.46
United Kingdom	−15.11	−0.63	23.58	0.82	−36.58	−1.74	−3.59	−0.14	−31.31	−1.43
United States	12.12	0.44	39.91	1.30	−30.39	−1.38	−10.16	−0.41	−19.86	−0.85
No HDI category										
North Korea	122.29	3.12	100.21	2.71	24.60	0.85	16.06	0.57	11.03	0.40
Puerto Rico	53.08	1.65	99.01	2.68	−3.07	−0.12	35.34	1.17	−23.08	−1.00

Countries were categorised into four groups as per HDI value in 2015: very high (HDI > 0.800), high (0.700 < HDI < 0.799), medium (0.550 < HDI < 0.669) and low (HDI < 0.550). Source: Authors' calculation from estimates of Global burden of disease 2016 study [12]

ASIR Age-standardised incidence rate, *ASMR* age-standardised mortality rate, *MIR* mortality-to-incidence ratio, *APC* annual percent change

Table 3 Country-wise human development index and its components in 2015

HDI rank	Country	Human Development Index (HDI)	Life expectancy at birth (years)	Expected years of schooling (years)	Mean years of schooling (years)	Gross national income (GNI) per capita (in 2011 PPP $)
Low HDI						
169	Afghanistan	0.479	60.7	10.1	3.6	1871
150	Angola	0.533	52.7	11.4	5.0	6291
153	Cameroon	0.518	56.0	10.4	6.1	2894
176	Democratic Republic of Congo	0.435	59.1	9.8	6.1	680
171	Côte d'Ivoire	0.474	51.9	8.9	5.0	3163
174	Ethiopia	0.448	64.6	8.4	2.6	1523
163	Haiti	0.493	63.1	9.1	5.2	1657
158	Madagascar	0.512	65.5	10.3	6.1	1320
152	Nigeria	0.527	53.1	10.0	6.0	5443
165	Sudan	0.490	63.7	7.2	3.5	3846
149	Syria	0.536	69.7	9.0	5.1	2441
151	Tanzania	0.531	65.5	8.9	5.8	2467
163	Uganda	0.493	59.2	10.0	5.7	1670
168	Yemen	0.482	64.1	9.0	3.0	2300
154	Zimbabwe	0.516	59.2	10.3	7.7	1588
Medium HDI						
139	Bangladesh	0.579	72.0	10.2	5.2	3341
118	Bolivia	0.674	68.7	13.8	8.2	6155
143	Cambodia	0.563	68.8	10.9	4.7	3095
111	Egypt	0.691	71.3	13.1	7.1	10,064
139	Ghana	0.579	61.5	11.5	6.9	3839
131	India	0.624	68.3	11.7	6.3	5663
113	Indonesia	0.689	69.1	12.9	7.9	10,053
121	Iraq	0.649	69.6	10.1	6.6	11,608
146	Kenya	0.555	62.2	11.1	6.3	2881
107	Moldova	0.699	71.7	11.8	11.9	5026
123	Morocco	0.647	74.3	12.1	5.0	7195
145	Myanmar	0.556	66.1	9.1	4.7	4943
144	Nepal	0.558	70.0	12.2	4.1	2337
147	Pakistan	0.550	66.4	8.1	5.1	5031
110	Paraguay	0.693	73.0	12.3	8.1	8182
116	Philippines	0.682	68.3	11.7	9.3	8395
119	South Africa	0.666	57.7	13.0	10.3	12,087
115	Viet Nam	0.683	75.9	12.6	8.0	5335
High HDI						
83	Algeria	0.745	75.0	14.4	7.8	13,533
84	Armenia	0.743	74.9	12.7	11.3	8189
78	Azerbaijan	0.759	70.9	12.7	11.2	16,413
52	Belarus	0.796	71.5	15.7	12.0	15,629
81	Bosnia and Herzegovina	0.750	76.6	14.2	9.0	10,091
79	Brazil	0.754	74.7	15.2	7.8	14,145
56	Bulgaria	0.794	74.3	15.0	10.8	16,261
90	China	0.738	76.0	13.5	7.6	13,345
95	Colombia	0.727	74.2	13.6	7.6	12,762
66	Costa Rica	0.776	79.6	14.2	8.7	14,006
68	Cuba	0.775	79.6	13.9	11.8	7455
99	Dominican Republic	0.722	73.7	13.2	7.7	12,756
89	Ecuador	0.739	76.1	14.0	8.3	10,536
70	Georgia	0.769	75.0	13.9	12.2	8856
69	Iran	0.774	75.6	14.8	8.8	16,395
86	Jordan	0.741	74.2	13.1	10.1	10,111
56	Kazakhstan	0.794	69.6	15.0	11.7	22,093
76	Lebanon	0.763	79.5	13.3	8.6	13,312
102	Libya	0.716	71.8	13.4	7.3	14,303
59	Malaysia	0.789	74.9	13.1	10.1	24,620
77	Mexico	0.762	77.0	13.3	8.6	16,383
87	Peru	0.740	74.8	13.4	9.0	11,295
66	Serbia	0.776	75.0	14.4	10.8	12,202
73	Sri Lanka	0.766	75.0	14.0	10.9	10,789
87	Thailand	0.740	74.6	13.6	7.9	14,519
97	Tunisia	0.725	75.0	14.6	7.1	10,249
71	Turkey	0.767	75.5	14.6	7.9	18,705
84	Ukraine	0.743	71.1	15.3	11.3	7361
54	Uruguay	0.795	77.4	15.5	8.6	19,148
105	Uzbekistan	0.701	69.4	12.2	12.0	5748
71	Venezuela	0.767	74.4	14.3	9.4	15,129
Very high HDI						
45	Argentina	0.827	76.5	17.3	9.9	20,945
2	Australia	0.939	82.5	20.4	13.2	42,822
24	Austria	0.893	81.6	15.9	11.3	43,609
22	Belgium	0.896	81.0	16.6	11.4	41,243
10	Canada	0.920	82.2	16.3	13.1	42,582
38	Chile	0.847	82.0	16.3	9.9	21,665
45	Croatia	0.827	77.5	15.3	11.2	20,291
28	Czech Republic	0.878	78.8	16.8	12.3	28,144
5	Denmark	0.925	80.4	19.2	12.7	44,519
23	Finland	0.895	81.0	17.0	11.2	38,868
21	France	0.897	82.4	16.3	11.6	38,085
4	Germany	0.926	81.1	17.1	13.2	45,000
29	Greece	0.866	81.1	17.2	10.5	24,808
43	Hungary	0.836	75.3	15.6	12.0	23,394
8	Ireland	0.923	81.1	18.6	12.3	43,798
19	Israel	0.899	82.6	16.0	12.8	31,215
26	Italy	0.887	83.3	16.3	10.9	33,573
17	Japan	0.903	83.7	15.3	12.5	37,268
18	South Korea	0.901	82.1	16.6	12.2	34,541
37	Lithuania	0.848	73.5	16.5	12.7	26,006
7	Netherlands	0.924	81.7	18.1	11.9	46,326
13	New Zealand	0.915	82.0	19.2	12.5	32,870
1	Norway	0.949	81.7	17.7	12.7	67,614
36	Poland	0.855	77.6	16.4	11.9	24,117
41	Portugal	0.843	81.2	16.6	8.9	26,104
50	Romania	0.802	74.8	14.7	10.8	19,428
49	Russia	0.804	70.3	15.0	12.0	23,286

Breast Cancer

Springer

Table 3 (continued)

HDI rank	Country	Human Development Index (HDI)	Life expectancy at birth (years)	Expected years of schooling (years)	Mean years of schooling (years)	Gross national income (GNI) per capita (in 2011 PPP $)
38	Saudi Arabia	0.847	74.4	16.1	9.6	51,320
5	Singapore	0.925	83.2	15.4	11.6	78,162
40	Slovakia	0.845	76.4	15.0	12.2	26,764
25	Slovenia	0.890	80.6	17.3	12.1	28,664
27	Spain	0.884	82.8	17.7	9.8	32,779
14	Sweden	0.913	82.3	16.1	12.3	46,251
2	Switzerland	0.939	83.1	16.0	13.4	56,364
16	UK	0.909	80.8	16.3	13.3	37,931
10	USA	0.920	79.2	16.5	13.2	53,245

Countries were categorised into four groups as per HDI value in 2015: very high (HDI > 0.800), high (0.700 < HDI < 0.799), medium (0.550 < HDI < 0.669) and low (HDI < 0.550). Data Source: United Nations Development Program [13]

References

1. GBD 2016 Cause of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2016;390(10100):1151–210.
2. Global Burden of Disease Cancer Collaborators. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990–2016. a systematic analysis for the Global Burden of Disease Study. JAMA Oncol. 2018. https://doi.org/10.1001/jamaoncol.2018.2706.
3. Sankaranarayanan R, Swaminathan R, Brenner H, Chen K, Chia KS, Chen JG, et al. Cancer survival in Africa, Asia, and Central America: a population-based study. Lancet Oncol. 2010;11(2):165–73.
4. Vostakolaei F, Karim-Kos HE, Janssen-Heijnen ML, Visser O, Verbeek AL, Kiemeney LA. The validity of the mortality to incidence ratio as a proxy for site-specific cancer survival. Eur J Public Health. 2010;21(5):573–7.
5. Chen SL, Wang SC, Ho CJ, Kao YL, Hsieh TY, Chen WJ, et al. Prostate cancer mortality-to-incidence ratios are associated with cancer care disparities in 35 countries. Sci Rep. 2017;7:40003.
6. Wang SC, Sung WW, Kao YL, Hsieh TY, Chen WJ, Chen SL, et al. The gender difference and mortality-to-incidence ratio relate to health care disparities in bladder cancer: National estimates from 33 countries. Sci Rep. 2017;7(1):4360.
7. Tsai MC, Wang CC, Lee HL, Peng CM, Yang TW, Chen HY, et al. Health disparities are associated with gastric cancer mortality-to-incidence ratios in 57 countries. World J Gastroenterol. 2017;23(44):7881.
8. Wang CC, Tsai MC, Peng CM, Lee HL, Chen HY, Yang TW, et al. Favorable liver cancer mortality-to-incidence ratios of countries with high health expenditure. Eur J Gastroenterol Hepatol. 2017;29(12):1397–401.
9. Sunkara V, Hebert JR. The colorectal cancer mortality-to-incidence ratio as an indicator of global cancer screening and care. Cancer. 2015;121(10):1563–9.
10. Ginsburg O, Bray F, Coleman MP, Vanderpuye V, Eniu A, Kotha SR, et al. The global burden of women's cancers: a grand challenge in global health. Lancet. 2017;389(10071):847–60.
11. Foreman KJ, Lozano R, Lopez AD, Murray CJ. Modeling causes of death: an integrated approach using CODEm. Popul Health Metr. 2012;10(1):1.
12. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2016. (GBD 2016). 2017. Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME). http://ghdx.healthdata.org/gbd-results-tool. Accessed 5 July 2018.
13. United Nations Development Program (UNDP). Human Development Database. 2018. http://hdr.undp.org/en/data#. Accessed 30 June 2018 and 1 July 2018.
14. Porter P. "Westernizing" women's risks? Breast cancer in lower-income countries. N Engl J Med. 2008;358(3):213–6.
15. Arnold M, Pandeya N, Byrnes G, Renehan AG, Stevens GA, Ezzati M, et al. Global burden of cancer attributable to high body-mass index in 2012: a population-based study. Lancet Oncol. 2015;16(1):36–46.
16. Park SK, Kim Y, Kang D, et al. Risk factors and control strategies for the rapidly rising rate of breast cancer in Korea. J Breast Cancer. 2011;14:79–87.
17. Hortobagyi GN, de la Garza Salazar J, Pritchard K, Amadori D, Haidinger R, Hudis CA, et al. The global breast cancer burden: variations in epidemiology and survival. Clin Breast Cancer. 2005;6(5):391–401.
18. Jemal A, Siegel R, Ward E, Murray T, Xu J, Smigal C, et al. Cancer statistics, 2006. CA Cancer J Clin. 2006;56(2):106–30.
19. WHO Guide for effective programmes; cancer control: knowledge into action; module 3: early detection (free full text). 2018. http://www.who.int/cancer/publications/cancer_control_detection/en/. Accessed 6 Aug 2018.
20. Jørgensen KJ, Zahl PH, Gøtzsche PC. Overdiagnosis in organised mammography screening in Denmark. A comparative study. BMC Womens Health. 2009;9(1):36.
21. Gøtzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2009. https://doi.org/10.1002/14651858.CD001877.pub3.
22. Jørgensen KJ, Zahl PH, Gøtzsche PC. Breast cancer mortality in organised mammography screening in Denmark: comparative study. BMJ. 2010;340:c1241.
23. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. N Engl J Med. 2012;367(21):1998–2005.
24. Kalager M, Adami HO, Bretthauer M, Tamimi RM. Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. Ann Intern Med. 2012;156(7):491–9.
25. Corbex M, Burton R, Sancho-Garnier H. Breast cancer early detection methods for low and middle-income countries, a review of the evidence. Breast. 2012;21(4):428–34.
26. Sankaranarayanan R, Ramadas K, Thara S, Muwonge R, Prabhakar J, Augustine P, Venugopal M, Anju G, Mathew BS. Clinical breast examination: preliminary results from a cluster randomized controlled trial in India. J Natl Cancer Inst. 2011;103(19):1476–80.
27. Luthar UK. Clinical downstaging of cancer of the uterine cervix e an interim strategy for developing countries. In: Proceedings of the UICC congress. 1994.
28. Devi BCR, Tang TS, Corbex M. Reducing by half the percentage of late-stage presentation for breast and cervix cancer over 4 years: a pilot study of clinical downstaging in Sarawak, Malaysia. Ann Onc. 2007;18(7):1172–6.
29. World Bank Database. 2018. WDI data. http://databank.worldbank.org/data. Accessed 25 July 2018.
30. Hogan DR, Stevens GA, Hosseinpoor AR, Boerma T. Monitoring universal health coverage within the Sustainable Development Goals: development and baseline data for an index of essential health services. Lancet Glob Health. 2018;6(2):e152–68.

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

Breast Cancer

Springer

International Journal of Research in Engineering and Innovation (IJREI)
journal home page: http://www.ijrei.com
ISSN (Online): 2456-6934

Comparison of half effect absorption-compression cascaded refrigeration system using thermodynamic (energy-exergy) analysis

R.S. Mishra

Department of Mechanical & Production Engineering, Delhi Technological University Delhi, India

Abstract

In the most of industrial processes, the waste heat energy increases production cost. Using this waste heat to produce power or cooling can decrease the costs and increases the thermo-economic efficiency of the industrial systems. A growing need for refrigeration in industry due to demands of higher living standards, the increasing requirements for comfort, and the increasing thermal load a half effect, single & double effect and triple effect LiBr-H_2O vapour absorption refrigeration systems provide very good efficiency for cooling up to 0^0C temperature . The energy-exergy analysis in the vapour absorption and compression systems helped to find out the major loss incurring component. In order to reduce those losses many attempts has been made so that first law efficiency (COP) of the vapour absorption refrigeration system cascaded with vapour compression cycle can be increased. It is found that the thermodynamic performances in the case of cascaded half effect vapour absorption refrigeration system coupled with vapour compression cycle is improved by 44.65% increment of first law efficiency (i.e. over all COP), 172.867% increment of second law efficiency (i.e. exergetic efficiency) of the half effect vapour absorption refrigeration cascaded with vapour compression cycle using HFC-134a, 42.867% enhancement in first law efficiency (COP) of 142.7267% increment of second law efficiency using HFO -1234yf for -50ºC of evaporator temperature of VCRS . Similarly 72.024% reduction in exergy destruction ratio based on exergy of output of the half effect vapour absorption refrigeration cascaded with vapour compression cycle using HFC-134a and 70.44% reduction in exergy destruction ratio using HFO-1234yf ecofriendly refrigerant for -50ºC of evaporator temperature of VCRS. The performances of single effect cascaded vapour absorption refrigeration system coupled with vapour compression cycle significantly higher than cascaded half effect vapour absorption refrigeration coupled with vapour compression cycle.
© 2019 ijrei.com. All rights reserved

Keywords: Energy-exergy analysis, Absorption-compression cascaded refrigeration system, Thermodynamic analysis

1. Introduction

For the rapid growth and development of the countries there has been continues demand of energies. These energies are generally meet with burning of fossil fuels which has led to global warming. So for meeting the huge demand of Refrigeration and air conditioning at large scale Vapor absorption system working on Lithium Bromide and water is used. Solar energy and Waste heat from industries are the most favorable options because these sources do not degrade the environment anyway.

The research work done on VAS working on LiBr/H2O for cooling purpose is based on First and Second Law of thermodynamics has been done in context of one of its major component Generator. The researchers have tried to find out the optimum COP of the system varying the different parameters affecting its performance. They found that Generator temperature has highest influence on the COP of the system.

1.1 Energy-exergy analysis of vapour absorption system

Energy analysis of vapour absorption refrigeration system and its effect on Generator The basic thermodynamic analysis is the Energy analysis for any system. Energy analysis helps to find out the major loss incurring component. In order to reduce

Corresponding author: R. S. Mishra
Email Address: hod.mechanical.rsm@dtu.ac.in

those losses many attempts has been made so that COP of the cycle can be increased. Energy from external source is given in Generator so generator temperature is found to play an important role in COP of the cycle. The second laws, have been employed to be a tool for consideration of the quantity and quality of energy in the thermal system. Exergy analysis has determined the destruction and losses of exergy in various components and hence exergy destruction ratio which is calculated based on exergy of overall system input or exergy of overall system output. Refrigeration systems are generally driven by electrical power in the mechanical vapour compression system. Absorption chillers are an alternative approach; whose use the thermal power (industrial waste heat, renewable energy sources, or other thermal sources) as a driving force. Furthermore, absorption cooling systems do not cause ozone depletion, as long they use natural substances, as working fluids .A lot of researchers have studied various absorption refrigeration systems which use different thermal energy sources.

A parametric study of double effect series flow and double effect parallel flow vapour absorption machines is carried out by Meraj , Adnan Hafiz , M. Jamil Ahmad [1] to provide design engineers, appropriate information for developing sustainable machines to reduce irreversibility in each component of the whole system, thereby reducing net exergy loss and efforts have been made by performed Second law analysis and found that the exergetic efficiency first increases , reaches a maxima and then starts decreasing with increase in high pressure generator temperature and also concluded that the exergetic efficiency is more in parallel flow as compared to the series flow for the same generator temperatures in the vapour absorption refrigeration system. Similarly irreversibility first decreases in both cases, reaches a minimum and then increases. Irreversibility is less in parallel configuration as compared to the series one. Absorption cooling cycles are environmental and can use solar or waste heat for cooling with very small electric power.

SC Kaushik, Akhilesh Arora,[1] carried out the energy and exergy analysis of single effect and series flow double effect water–lithium bromide absorption systems by developing computational model using water–lithium bromide solution. This analysis involved the determination of effects of generator, absorber and evaporator temperatures on the energetic and exergetic performance of these absorption refrigeration systems. The effects of pressure drop between evaporator and absorber, and effectiveness of heat exchangers are also investigated for vapour absorption systems. The numerical computation was carried out for finding performance parameters (i.e. coefficient of performance, exergy destruction, efficiency defects and exergetic efficiency) and it is found that the coefficient of performance of the single effect system lies in range of 0.6–0.75 and 0.90 to 1.2 for double effect system respectively. Mohd. Meraj et.al [2] studied double effect series flow and double effect parallel flow vapour absorption machines and carried out second law analysis to reduce irreversibility in each component of the whole system, thereby reducing net exergy loss and found that the exergetic efficiency first increases , reaches a maxima and then starts decreasing with increase in high pressure generator temperature. . It has also been observed that the exergetic efficiency is more in parallel flow as compared to the series flow for the same generator temperatures Rabi Karaali [3] carried out work on thermal analysis of a double effect parallel flow and single effect absorption cooling systems by developing computer program for the thermodynamic properties of lithium bromide-water solutions using the exergy analysis and found that the double effect parallel flow absorption systems have better advantages than the single effect absorption system. The coefficient performance (COP) and the exergetic efficiency (exergetic coefficient of performance) of the double effect parallel flow absorption systems (i.e. the double effect cycle COP is 1.195 and exergetic efficiency as 0.28) , are higher than the single effect cycle (COP is 0.68) and exergetic efficiency (as 0.23) . Rabi Karaali [4] presented exergy analysis of a half effect absorption system using thermodynamic properties of lithium bromide-water solutions, by developing computer program in FORTRAN codes. Due to the reason that the half effect absorption systems is used for low temperature heat energy or waste heat energy and found that the coefficient performance (COP) and the exergetic efficiency of the half effect absorption system are 0.45 and 0.24, respectively. For each component the exergy loss is calculated and it was found that the 92 % of total irreversibilities in the absorber and in the evaporator and also concluded that better design and efficiencies of the evaporator and the absorber is very important for the half effect absorption cycles. By improving the design of these two components will directly improve and affect the working conditions and thermal performances of the cycle. Md Azhar and M. Altamush Siddiqui [5] analyzed Triple Effect Vapour Absorption cycle using liquefied petroleum gas (LPG) and compressed natural gas (CNG) as sources of energy selected because the triple effect system requires heat at relatively high temperature using LiBr-H2O solution as working pair and compared with the Single and Double Effect cycles. The thermodynamic analysis is done for different values of the evaporator and the condenser temperatures; the absorber temperature is assumed to be equal to the main condenser temperature by assuming the temperature and concentration of LiBr salt in the high pressure generator, to which heat is supplied, are varied simultaneously for fixed temperatures/pressures in the middle and high pressure condensers and investigated the effect of the high pressure generator temperature on the coefficient of performance (COP), volume flow rate and cost of the LPG and CNG. Arora et al. [6] have developed a double effect cycle 120-150^{0}C, half effect 60-75^{0}C and triple effect cycles which require generator temperature even higher than 150^{0}C.

Dixit Manoj et.al [7] developed mathematical model of vapour absorption-compression cascaded refrigeration system, comprising of a vapour compression refrigeration system in low temperature stage using CO2, NH3 and R134a as ecofriendly refrigerants and H_2O-LiBr vapour absorption

refrigeration system at the high temperature stage and found that the cascade condenser, compressor and refrigerant throttle valve are the major source of exergy destruction**.** S Arivazhagan et.al. [8] carried out experimental investigation 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). The performance of the system in terms of degassing range, coefficient of performance and second law efficiency has been obtained. The system is capable of producing evaporating temperature as low as -7^0C with generator temperatures ranging from 55^0 to 75^0C. The degassing range is 40% more in high absorber than in low absorber as the high absorber is operated at optimum intermediate pressure. The optimum generator temperature is found to be in the range of 65–70^0C at which the coefficient of performance is 0.36. Shun-Fu Lee and S. A. Sheriff [9] carried out parametric analysis using thermodynamic analysis of a lithium bromide/water absorption system for cooling and heating applications and it was found that a low cooling water temperature yields both a higher cooling COP and higher exergetic efficiency and also observed that by increasing the heat source generator temperature can improve the COP of the system and also found that the heat source temperature increases beyond a certain limit , the COP of the system decreases due to the negative effect of increasing the heat source temperature. Inzunza et al. [10], compared of the performance of single-effect, half-effect, double-effect in series and inverse absorption cooling systems operating with the mixture LiBr- H_2O and found that for the generation temperature between 100^0C and 110^0C, the COP of the single effect was up to 0.89, for the generation temperature of over 55 ^{0}C the COP of the half effect was up to 0.44. They also found that the most efficient system is double effect of the COP up to 1.48 for higher temperature while, for low temperatures the half effect systems work better than any other. Wang and Zheng [11] developed the performance of half effect & single effect LiBr / H_2O absorption cooling cycle and found that half effect cycle's absorption cooling cycle can be used for generator temperature from 50^0C to 80^0C. This paper mainly deals with half effect vapour absorption-compression cascade system is lowest by using R141b and highest while using R227ea as shown in table-2. The cascade system exergetic refrigeration system, comprising of a vapour compression cascade refrigeration system in low temperature stage and half effect absorption refrigeration system at the high temperature stage, is analyzed. CO_2, NH3 and R134a have been considered as refrigerants in the compression stage and the H2O-LiBr refrigerant absorbent pair in the absorption stage using energy-exergy analysis. The performance of the refrigeration system. The comparative study carried out in terms of their thermal performances using HFC-134a and HFo-1234yf for evaporator temperature of VCRS is -50^0C and -30^0C respectively using HFO-1234ze respectively.

2. Results and Discussion

Following numerical values have been taken for validation of thermal models for half effect LiBr-H_2O cascade vapour compression-absorption refrigeration systems.

2.1 Half Effect Cascade vapour compression-absorption refrigeration system using ecofriendly refrigerants

$T_{_Generator}$=80ºC.T_Cond=30ºC, $T_{_Eva}$=5ºC,
T_absorber=30ºC, Compressor Efficiency=0.80,
T_Eva_LTC=-50ºC (for using R-1234yf and HFCs, hydrocarbons, natural refrigerants in VCRS), & -30ºC (for using R1234ze & R1234yf, HFCs, hydrocarbons, natural refrigerants in the VCRS), approach (Temperature Overlapping) =10ºC)

Table-1 shows the variation of thermal performance parameters of triple effect cascade vapour absorption refrigeration system using lithium bromide -water solution coupled with vapour compression refrigeration system using 29.09 kW of load. The evaporator of VARS is cascaded with condenser of vapour compression refrigeration systems using following nine ecofriendly refrigerants. It is found that overall first law efficiency in terms of coefficient of performance (COP) of cascade refrigeration system using R245fa refrigerant is highest and lowest by using R227ea as shown inTable-1(a) and Table-1(b) respectively while exergy destruction of
efficiency is highest by using R141b and lowest by using R227ea as shown in table-3.

Table-1(a):Thermal performance of single Effect LiBrH_2O Vapour absorption refrigeration system cascaded by vapour compression refrigeration system using ecofriendly refrigerants at the evaporator temperature of -30ºC in the VCRS cycle

S.No	Refrigerant	COP_Over_alll	EDR_Overall	Exergetic efficiency
1	R134a	1.126	2.057	0.3271
2	R1234ze	1.124	2.070	0.3257
3	R1234yf	1.117	2.122	0.3203
4	R227ea	1.117	2.218	0.3107
5	R236fa	1.119	2.108	0.3217
6	R245fa	1.133	2.008	0.3324
7	R32	1.122	2.087	0.3239
8	R507a	1.110	2.174	0.3151
9	R717	1.129	2.04	0.3289

10	R123	1.138	1.976	0.3360
11	R125	1.099	2.260	0.3068
12	R152a	1.134	2.003	0.3330
13	R404a	1.105	2.216	0.3109
14	R410a	1.122	2.086	0.3240
15	R407c	1.004	2.382	0.2957
16	R290	1.125	2.069	0.3258
17	R600	1.133	2.011	0.3321
18	R600a	1.125	2.063	0.3265
19	R141b	1.145	1.929	0.3414
20	R143a	1.141	2.154	0.3171

Table-1(b)::Thermal performance of Half Effect LiBrVapour absorption refrigeration system cascaded by vapour compression refrigeration system using ecofriendly refrigerants (for T_EVA_VCRS= -50ºC)

S.No	Refrigerant	COP_Over_alll	EDR_Overall	Exergetic Efficiency
1	R134a	0.6155	1.516	0.3974
2	R1234yf	0.6079	1.602	0.3844
3	R227ea	0.5956	1.746	0.3642
4	R236fa	0.6084	1.596	0.3853
5	R245fa	0.6196	1.473	0.4044
6	R32	0.6138	1.535	0.3945
7	R507a	0.6042	1.644	0.3782
8	R717	0.6157	1.515	0.3977
9	R123	0.6238	1.428	0.4118
10	R125	0.5951	1.752	0.3634
11	R404a	0.6011	1.680	0.3732
12	R410a	0.6142	1.532	0.3950
13	R407c	0.5956	1.746	0.3641
14	R290	0.6151	1.521	0.3967
15	R600	0.6203	1.465	0.4057
16	R600a	0.6144	1.528	0.3955
17	R141b	0.6296	1.368	0.4223
18	R143a	0.6068	1.614	0.3825

Table-2:Thermal performance of double Effect LiBrH_2O Vapour absorption refrigeration system cascaded by vapour compression refrigeration system using ecofriendly refrigerants at the evaporator temperature of -30ºC in the VCRS cycle,

S.No	Refrigerant	COP_Over_alll	EDR_Overall	Exergetic Efficiency
1	R134a	1.822	1.037	0.4910
2	R1234ze	1.817	1.047	0.4884
3	R1234yf	1.80	1.085	0.4796
4	R227ea	1.77	1.156	0.4638
5	R236fa	1.805	1.075	0.4820
6	R245fa	1.838	1.001	0.4998
7	R32	1.813	1.057	0.4863
8	R507a	1.784	1.123	0.4711
9	R717	1.829	1.021	0.4949
10	R123	1.849	0.9767	0.5059
11	R125	1.757	1.187	0.4573
12	R152a	1.84	0.9964	0.5009
13	R404a	1.772	1.152	0.4647
14	R410a	1.812	1.059	0.4858
15	R407c	1.731	1.251	0.4442
16	R290	1.818	1.045	0.4889
17	R600	1.837	1.003	0.4994
18	R600a	1.819	1.041	0.4899
19	R141b	1.865	0.9424	0.5148
20	R143a	1.79	1.108	0.4744

Table-3:Thermal performance of Tripple Effect LiBrH_2O Vapour absorption refrigeration system cascaded by vapour compression refrigeration system using ecofriendly refrigerants at the evaporator temperature of -30ºC in the VCRS cycle,

S.No,1	Refrigerant	COP$_{_Over_alll}$	EDR$_{Overall}$	Exergetic Efficiency
1	R134a	2.151	1.421	0.4131
2	R1234ze	2.148	1.429	0.4116
3	R1234yf	2.131	1.469	0.4050
4	R227ea	2.10	1.543	0.3933
5	R236fa	2.135	1.459	0.4067
6	R245fa	2.168	1.382	0.4198
7	R32	2.139	1.448	0.4084
8	R507a	2.113	1.510	0.3984
9	R717	2.154	1.414	0.4142
10	R123	2.179	1.357	0.4242
11	R125	2.087	1.574	0.3885
12	R152a	2.169	1.380	0.4202
13	R404a	2.098	1.547	0.3926
14	R410a	2.142	1.442	0.4095
15	R407c	2.034	1.712	0.3687
16	R744	1.993	1.827	0.3537
17	R290	2.147	1.430	0.4115
18	R600	2.167	1.385	0.4194
19	R600a	2.148	1.425	0.4124
20	R141b	2.196	1.321	0.4309
21	R143a	2.12	1.495	0.4008

Table 4(a): Performance comparison of single effect vapour absorption refrigerated cascaded with vapour compression refrigeration system using HFC-134a at evaporator temperature of -50ºC .

Performance Parameters	Half effect Vapour Absorption System	Half effect Cascade VARS_VCRS	% improvement
COP	0.4255	0.6155	44.65
Exergetic Efficiency	0.1558	0.3974	155
EDR	5.419	1.516	72.022 (Reduction)

The performance parameters have been compared with the half effect vapour absorption system with half effect vapour absorption-compression refrigeration system cascaded with vapour compression refrigeration system using ecofriendly HFC-134a and it was observed that 44.65% first law efficiency enhancement and 42.876% using HFO-1234yf for -50ºC of evaporator temperature of vapour compression cycle as shown in table-4(a) & table-4(b) respectively. Similarly 155% enhancement in second law efficiency using half effect vapour absorption system cascaded with vapour compression refrigeration cycle using HFC-134a and 146.726% using HFO-1234yf in the vapour compression cycle as shown in table-4(a-b) respectively. Similarly exergy reduction is 72.024% using HFC-134a and 146.726% using HFO-1234yf in the vapour compression cycle as shown in table-4(a) & table-4(b) respectively. The performance parameters have been compared with the half effect vapour absorption system with half effect vapour absorption-compression refrigeration system cascaded with vapour compression refrigeration system using ecofriendly HFC-134a, single effect vapour absorption system with half effect vapour absorption-compression refrigeration system cascaded with vapour compression refrigeration system using ecofriendly HFC-134a, double effect vapour absorption system with half effect vapour absorption-compression refrigeration system cascaded with vapour compression refrigeration system using ecofriendly HFC-134a, tripplel effect vapour absorption system with half effect vapour absorption-compression refrigeration system cascaded with vapour compression refrigeration system using ecofriendly HFC-134a respectively and it is found that first law efficiency (COP) and second law efficiency (exergetic efficiency) performance increases while exergy destruction based on exergy of product is decreases

.

Table 4(b) : Performance comparison of single effect vapour absorption refrigerated cascaded with vapour compression refrigeration system using HFO-1234yf at evaporator temperature of -50ºC .

Performance Parameters	Half effect Vapour Absorption System	Half effect Cascade VARS_VCRS	% improvement
COP	0.4255	0.6079	42.876
Exergetic Efficiency	0.1558	0.3844	146.726
EDR	5.419	1.602	70.44

Table-5 : Performance comparison of single effect vapour absorption refrigerated cascaded with vapour compression refrigeration system using HFO-1234yf at evaporator temperature of -50ºC .

Performance Parameters	Half effect simple vapour absorptin refrigeration system	Half effect cascaded vapour Absorption System	Single effect Cascaded vapour absorption System	Double Effect cascaded vapour absorption System	Triple effect cascaded vapour absorption System
COP	0.4255	0.6155	1.126	1.822	2.151
Exergetic Efficiency	0.1558	0.3974	0.3271	0.4131	0.4910
EDR	5.519	2.057	1.516	1.421	1.037

3. Conclusions

The following conclusions were drawn.

(i) The thermodynamic performances in the case of cascaded half effect vapour absorption refrigeration system coupled with vapour compression cycle is improved.

(ii) The performances of single effect cascaded vapour absorption refrigeration system coupled with vapour compression cycle significantly higher than cascaded half effect vapour absorption refrigeration coupled with vapour compression cycle.

(iii) The 44.65% increment of first law efficiency (i.e. over all COP) of the half effect vapour absorption refrigeration cascaded with vapour compression cycle using HFC-134a ecofriendly refrigerant for -50ºC of evaporator temperature of VCRS.

(iv) The 42.867% increment of first law efficiency (i.e. over all COP) of the half effect vapour absorption refrigeration cascaded with vapour compression cycle using HFC-134a ecofriendly refrigerant for -50ºC of evaporator temperature of VCRS.

(v) The 172.867% increment of second law efficiency (i.e. exergetic efficiency) of the half effect vapour absorption refrigeration cascaded with vapour compression cycle using HFC-134a ecofriendly refrigerant for -50ºC of evaporator temperature of VCRS.

(vi) The 142.7267% increment of second law efficiency (i.e. exergetic efficiency) of the half effect vapour absorption refrigeration cascaded with vapour compression cycle using HFC-134a ecofriendly refrigerant for -50ºC of evaporator temperature of VCRS.

(vii) The 72.024% reduction in exergy destruction ratio based on exergy of output of the half effect vapour absorption refrigeration cascaded with vapour compression cycle using HFC-134a ecofriendly refrigerant for -50ºC of evaporator temperature of VCRS.

(viii) The 70.44% reduction in exergy destruction ratio based on exergy of output of the half effect vapour absorption refrigeration cascaded with vapour compression cycle using HFC-134a ecofriendly refrigerant for -50ºC of evaporator temperature of VCRS.

References

[1] SC Kaushik, A Arora [2009] Energy and exergy analysis of single effect and series flow double effect water–lithium bromide absorption refrigeration systems, International journal of Refrigeration, Vol-32, No-6page-1247-1258

[2] Mohd. Meraj , Adnan Hafiz , M. Jamil Ahmad [2016] Second Law Analysis of Series and Parallel Flow Double Effect Vapor Absorption Chiller, IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE) e-ISSN: 2278-1684,p-ISSN: 2320-334X, Volume 13, Issue 3 Ver. VI (May- Jun. 2016), PP 104-110.

[3] Rabi Karaali [2017] Exergetic Comparison of Single and Double Effect Absorption Cooling Cycles,Karaelmas Fen ve Müh. Derg. 7(1):122-128, 2017.

[4] Rabi Karaali [2016] Exergy analysis of half-effect absorption cooling cycles, international journal of engineering sciences & research technology (IJESRT).ISSN: 2277-9655,Vol-5, issue-9 .page- 405-411.

[5] Md Azhar and M. Altamush Siddiqui[2013] Thermodynamic analysis of a gas operated triple effect absorption cycle, International Journal of Innovative Research in Science, Engineering and Technology , Vol. 2, ISSN: 2319-8753 , Issue 5.

[6] Akhilesh Arora,Manoj Dixit,S.C. Kaushik.[2018] : Computation of optimum parameters of a half effect water-lithium bromide vapour absorption refrigeration system, Journal of Thermal Engineering, Vol-2, No-2, page- 683-692.

[7] M Dixit, SC Kaushik, A Arora [2018] Energy and exergy analysis of absorption-compression cascade refrigeration system, Journal of Thermal Engineering , Vol-2 (6), page-995-1006.

[8] Arivazhagan S, et.al. [2006]Experimental studies on HFC based two-stage half effect vapour absorption cooling system, Applied Thermal Engineering, Vol-26, Issues-14-15, Pages 1455-1462.

[9] Lee Shun.Fu., Sherif, S.A. [2000] Thermodynamic analysis of a lithium bromide/water absorption system for cooling and heating applications, International Journal of energy Research,Vol.25, pp.1019-1031.

[10] L.A. Domínguez-Inzunza, J.A. Hernández-Magallanes, M. Sandoval-Reyes, and W. Rivera[2014], "Comparison of the performance of single-effect, half-effect, double-effect in series and inverse and triple-effect absorption cooling systems operating with the NH3-LiNO3 mixture" Applied Thermal Engineering VOL.66, pp.612- 620,.

[11] J, Wang and D. Zheng, [2009]"Performance of one and half effect absorption cooling cycle of H_2O/LiBr system", Energy Conversion and Management, VOL. 50, pp. 3087-3095.

Cite this article as: R.S. Mishra, Comparison of half effect absorption-compression cascaded refrigeration system using thermodynamic (energy-exergy) analysis, , *International journal of research in engineering and innovation (IJREI),* vol 3, issue 1 (2019), 6-11.

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-019-7234-z

Cyberbullying detection on social multimedia using soft computing techniques: a meta-analysis

Akshi Kumar[1] · Nitin Sachdeva[1]

Received: 26 September 2018 / Revised: 4 January 2019 / Accepted: 15 January 2019
Published online: 23 January 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract
Cyberbullying is to bully someone in the digital realm. It has become extremely detrimental as the social media and the internet have become more popular and omnipresent. People use the internet services to viciously attack others from behind a screen. The substantial growth in the dimensionality, heterogeneity, subjectivity and multimodality of social media and the pressing need to timely curtail the damage instigated through cyberbullying, has fostered the need to devise automated mechanisms which detect such unfavorable activities. The use of soft computing techniques to handle such pernicious issue has been studied invariably and widely in literature. This study is to understand the viability, scope and significance of this alliance of using soft computing techniques for cyberbullying detection on social multimedia. This work is a systematic literature review to gather, explore, comprehend and analyze the research trends, gaps and prospects of this pairing in a well-organized way. The contribution of this study is noteworthy as it focuses on the use and application of soft computing techniques for cyberbullying detection on social multimedia utilizing a meta-analytic approach in order to integrate, interpret and critically analyze the findings in the original studies for expounding novel approaches to achieve comparable and effectual results pertaining to the defined research domain. Published studies starting April 2003, accessed from six digital portals (ACM, IEEE, Elsevier, Wiley, Springer and Taylor and Francis) have been reviewed to expound the state-of-art within the domain to give insightsand finally identify the directions of future research.

Keywords Cyberbullying · Social multimedia · Soft computing · Machine learning · Meta-analysis

✉ Akshi Kumar
akshikumar@dce.ac.in

Nitin Sachdeva
nits.usit@gmail.com

[1] Department of Computer Science & Engineering, Delhi Technological University, Delhi, India

1 Introduction

Information is power, but without a means to distribute information, people cannot harness this power. Social media has emerged as a key player which provides a platform for expression and distribution of content in today's world.The primary intent of social networking sites is to create, professional, interest, relationship-based virtual communities enabling stronger connections with everyone around the world. With the proliferation of these sites (Twitter, Tumblr, Google+, Facebook, Instagram, Snapchat, YouTube etc.), the user can post and share all kinds of multimedia content (text, image, audio, video) in the social setting using Internet without much knowledge about the Web's client-server architecture and network topology. The social networking sites have given 'everyone a voice' but at the same time, we're drowning in abundance, complexity of choices and unfortunately, the misappropriation or misdirection of influence. Moreover, when lots of individuals come together and that too from different countries, communities, races, ethnicities, gender, and varied age-groups, there are bound to be conflicts, controversies, and intimidation vulnerabilities. That is, although social networking sites proffer numerous benefits as these facilitate participation and collaboration but on the flip side hate speech, social distrust, cyberbullying, identity theft, cyber-stalking and cascading of rumors and fake stories are some antithetical concerns associated with it. The pervasive reach of these sites has irrefutably triggered, contributed and exacerbated bullying.

As per the National Bullying Prevention Center, *'Every child on Facebook likely has a bullying story, whether as the victim, bully or as a witness'* [35]. Cyberbullying is defined as bullying an individual or a group of individuals using Internet, mobiles or any other electronic device by sending inappropriate textual or non-textual multimedia message in order to hurt or cause embarrassment [52]. The one who bullies is called as 'bully' and the other is said to be 'victim'. The term 'Cyberbullying' was coined by Canadian educator and anti-bullying activist Bill Belseyin the year 2003 [10]. It is the repeated exposure of the negative actions on the part of one or more individuals in order to inflict humiliation, harassment, discomfort or injury upon another through the use of electronic medium [17] like emails, chat rooms, instant messaging, cell phones or by posting videos, audios, images etc. Bullying has been a part of the human civilization history which involves hurting someone either by humiliating or harassing in any form, involving mental, verbal or physical damage. When this assault takes place on cyberspace, it is referred to as cyberbullying/ cyberharassment/ cybervictimization [30].

According to a recent study, nearly 43% of the teenagers in the United States are victims of cyberbullying [53]. It is more persistent way of bullying an individual in front of the entire online community especially within the social setting which can eventually lead to psychological, mental and emotional breakdown for the victim inculcating the sense of low self-esteem, low self-confidence, anger, depression, stress, loneliness, sadness, health degradation etc. [49]. Many such intense cases have tragically ended in self-injury or suicides, underlining the grave nature of this critical issue [30]. The following Fig. 1 presents an example of cyberbullying from twitter.

With the technological advancement, the social freedoms that the networking sites give and larger audience cyberbullying has spread manifolds affecting the individual not only limited to their workplace but has also children and young adults in their daily lives. Anonymity further allows bullies to be more aggressive and offensive due to the reduced chance of being detected and punished, making it critical to efficiently detect cyberbullying behavior in a real- time setting. This poses significant threat to the physical and mental health of the victims making it



Fig. 1 Example of cyberbullying

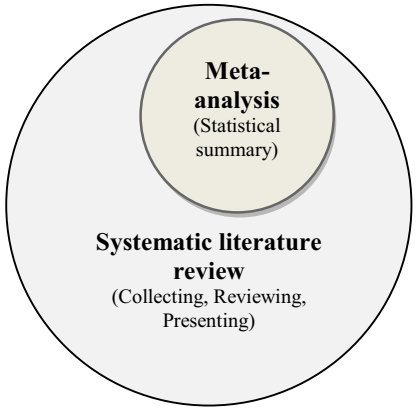
a public health concern. Various studies have reported that victims of cyberbullying have lower self-esteem, higher levels of depression, suffer from behavioral issues and addict to substance abuse. Bullying victimization may trigger a sequence of events that results in suicidal behavior. The first reported case of cyberbullying was of an American middle school student, Ryan Halligan of Vermont in 2003 [26]. Ryan was constantly bullied in person and online by his classmates and this bullying was attributed as his reason to commit suicide.

Pertinent studies indicate that social media is one of the mostly favored medium by bullies and various factors such as socio-demography, physiological distress and time frames are related to cyberbullying. The massive volumes of human-centric, real-time, multimodal, heterogeneous and unstructured social media data makes manual detection intractable. Moreover, the social web applications/services are not restricted to the text-based data but extend to the partially unknown complex structures of image, audio and video. This fosters the need to develop intelligent tools and techniques for identifying, detecting and assessing cyberbullying from the available social multimedia data to lower down its hazardous impact. Design and development of contemporary tools which tap and analyze online detrimental behavior automatically from the high-dimensional social multimedia are imperative. Automated cyberbullying detection is typically a classification problem where the intent is to classify each abusive or offensive comment/ post/ message/ image as either a bullying or a non-bullying.

Studies have been conducted to explore new paradigms which handle uncertainty, imprecision, approximation, and fuzziness generated by thesocial multimedia content making automated cyberbullying detection techniques computationally proficient. In recent years, soft computing based multimedia analytics as a technology-based solution for automated detection of cyberbullying has attracted a lot of attention by both researchers and practitioners. Soft computing is a blanket term which comprises of variety of techniques such as machine learning, fuzzy logic, swarm and evolutionary computing, deep learning, amongst others. This work is a study to understand the feasibility, scope and relevance of this alliance of using soft computing techniques for cyberbullying detection on social multimedia portals. A systematic literature review on the use of soft computing techniques to detect cyberbullying automatically from the social multimedia is presented. Systematic literature review (SLR) is a detailed,structured, criterion-based and formal approach of gathering and reviewing literature studies pertaining to well-defined research questions on the selected domain. Meta-analysis summarizes the results of these studies to generate a meta-summary of the qualitative filtered studies. It is a statistical pooling to quantify the systematic review and eventually build an evidence pyramid. A systematic review does not necessitate a meta-analysis, but a meta-analysis (if used to answer a quantitative question) must be based on a systematic review. The following Fig. 2 illustrates the relationship of SLR and meta-analysis.

Thus, a SLR may contribute to a quantitative evidence synthesis known as meta-analysis. Meta-analysis identifies heterogeneity among multiple studies, identifies data gap in the

Fig. 2 Relationship of SLR and meta-analysis



knowledge base and at the same time suggest directions for future research. It is also referred to as research synthesis, pooled analysis, or quantitative review. In this work, the observational studies are thus meta-analyzed to provide quantitative answers to the specific, pre-defined questions. The goal is to gather empirical evidences and analyze results from existing studies to give a critically evaluated discussion on the existing trends in available research, identify gaps in current search and provide future prospects in the area by means of answering the established research questions.

A previously published systematic survey on cyberbullying [44] focused on the approaches to automated detection, namely, supervised learning, lexicon based, rule based and mixed-initiative approaches. The survey conducted in this work maps the state-of-the-art in cyberbullying detection research focused essentially on automated approaches using soft computing. Published studies within past 15 years, starting April 2003, accessed from six main digital portals (ACM, IEEE, Elsevier, Wiley, Taylor and Francis and Springer) have been reviewed. A quantitative, scientific synthesis of research results known as meta-analysis is done to finally summarize the results. The rest of the paper is organized as follows: Section 2 explicates the two primary concepts discussed in this review, which are cyberbullying and soft computing. Section 3 elaborates the review methodology enlisting the research questions identified to conduct this study review. Section 4 overviews the literature survey of the selected studies concisely. Finally section 5 provides the results and discussion followed by the conclusion in section 6 which enlists the gaps for potential research efforts.

2 Automated cyberbullying detection and soft computing

Netiquette refers to good manners on the Internet and treating other people on the Internet as you would like to be treated yourself. Unfortunately, some people use the Internet and/or mobile phones to offend or harass others. This is referred to as cyberbullying. Automated cyberbullying detection is a pro-active strategic technology based tool. It is a generic classification task where the multimodal, multimedia content is categorized as bullying or non-bullying. It is characterized as a predictive learning model in the social setting which detects the presence of cyberbullying in an online post (textual/non-textual) so that it does not inflict

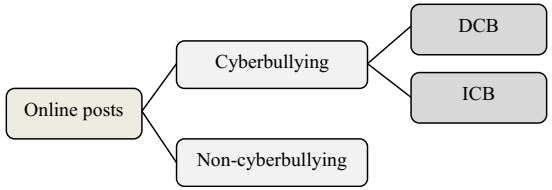


Fig. 3 Classification of social media posts

seriously or damage the victim's emotional, psychological and social state. The posts classified as bullying can further divided into two categories, namely, direct cyberbullying (DCB) and indirect cyberbullying (ICB) [45], as shown in Fig. 3. DCB involves direct sending of harmful content to a person either via email or SMS etc. ICB comprises of posting harmful contents about any person on social media or sharing it with others, for example posting an improper photograph of someone on Facebook is an example of ICB.

The direct and indirect cyberbullying is categorized into the different types as depicted in Fig. 4.

ICB messages are further categorized into six types [45]. These include Indirect Harassment, Denigration, Impersonation, Outing, Trickery and Exclusion, whereas DCB includes Flaming and Direct Harassment. The following Table 1 briefly explains these types of cyberbullying.

The elusive nature of cyberbullying undermines the self-esteem of the cyber victim, affecting him or her mentally, socially and psychologically. Automated detection model consists of multiple tasks which identify and classify posts as bullying or not. Considered as a generic classification problem, a typical cyberbullying detection process extracts the features from the pre-processed data and classifies the posts accordingly as shown in the Fig. 5 below.

The pre-processing phase includes cleaning the acquired data by removing unwanted URL's or strings etc., handling missing values, correcting words etc. and then transforming it in a representation suitable for feature extraction. After pre-processing, features such as keywords depicting bad/nasty/rude/abusive/hateful/attacking words, N-grams, pronouns, skip-grams are extracted. Next phase uses supervised learning techniques to classify the messages as either containing bullying content or not. This study comprehends the application of soft computing techniques to classify online bullying content. These techniques use approximate calculations to provide imprecise but usable solutions to complex computational problems. Also, referred to as computational intelligence techniques, soft computing techniques are generally divided into

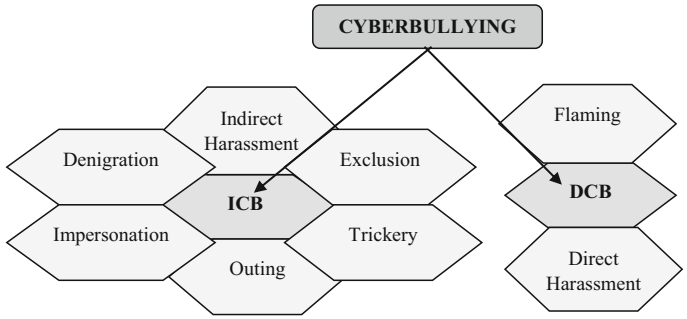


Fig. 4 Types of cyberbullying

Table 1 Types of cyberbullying

Types of cyberbullying	Details
Flaming	It is online fight between individuals. They usually exchange vulgar electronic messages [30, 45]. People fighting on online forums exchanging obscene message is an example of Flaming.
Direct harassment	It is directly harassing a person either by insulting or threatening him or her via messages [30, 45]. It includes only two parties- the one who bullies and the other who is bullied. Threatening or harassing a person either by sending email or SMS directly is an example of direct harassment.
Indirect harassment	It is indirectly harassing a person either by insulting or threatening him or her via messages posted online [30, 45]. It includes many parties. Posting embarrassing photos on social media in order to harass the other person indirectly is an example of indirect cyberbullying
Denigration	It is spreading hearsay or rumors about others in order to ruin their reputation [30, 45]. It puts the status of the cyber-victim on stake. Posting skewed contents on forums or blogs etc. in order to turn down cyber-victim's reputation is an example of denigration.
Impersonation	It is acting or pretending as other person and then doing anti-social activities in order to embarrass or damage his or her reputation [30, 45]. Imitating cyber-victim either by creating any fake profile or through hacking and sending messages that may instigate other users to attack the victim is an example of impersonation.
Outing	It is sharing private information of a cyber-victim without his or her consent in order to hurt the victim [30, 45]. Posting a humiliating picture of someone in order to hurt the cyber victim is an example of outing.
Trickery	It is obtaining sensitive information about a user by faking the trust of cyber victim and then eventually violating that trust [30, 45]. Obtaining a personal video by faking as close friend and then posting it online is an example of trickery.
Exclusion	It involves the exclusion of the cyber victim from online communities or groups etc. [30, 45]. Excluding a person knowingly from WhatsApp group is an example of exclusion.

following categories *(Machine learning (ML), Neural networks (NN), evolutionary computation, fuzzy logic and probabilistic reasoning)* [1] as depicted in the following Fig. 6.

The machine learning techniques are further categorized as shown in the Fig. 7 below.

From Figs. 6 and 7, it is apparent that soft computing is a *'blanket term'* comprising of several methodologies which are themselves inter-related to one other. Deep learning (DL) is considered as a part-of a broader family of ML [1] based on learning data representations, in contrast to the task-specific algorithms and where learning can be supervised, semi-supervised or unsupervised. DL consists of various techniques such as deep NN (DNN), recursive NN, recurrent NN, convolutional NN and deep belief networks, whereas neural networks is also an established sub-category of soft computing (SC) techniques [1] which includes feed forward; multi-layer perceptron; radial basis; Kohonen self-organizing; modular; shallow and deep NN (DNN). Thus, it can be inferred that SC, ML and DL are inter-connected to each other as shown in the following Fig. 8:

Social media is inherently an informal way of communication with all kinds of multimedia content. The following Fig. 9 depicts the multimedia types supported by popular social networking sites.

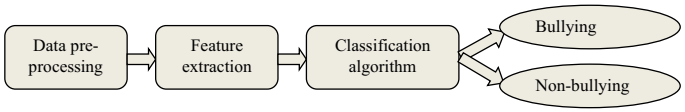


Fig. 5 Generic cyberbullying detection process

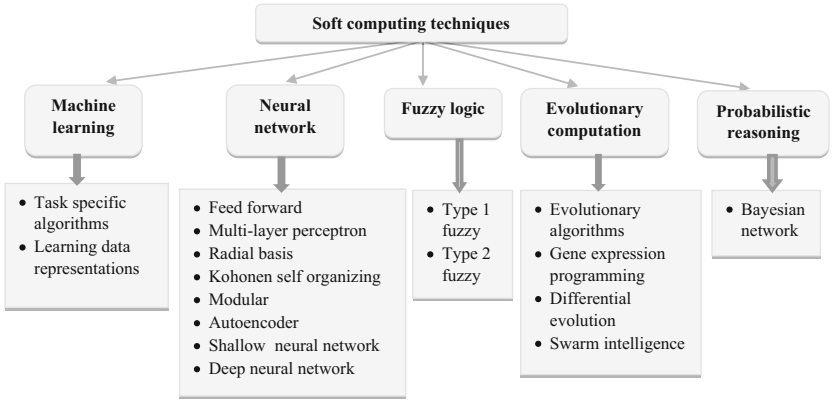


Fig. 6 Soft computing techniques

Social media dynamics keep changing with respect to increasing user base and user-activity which makes it a high -dimensional, complex and fuzzy data space for analytical processing. Soft computing has a solution to many real world problems. The guiding principle of soft computing is to exploit the tolerance for imprecision, uncertainty and partial truth to achieve robustness, low cost solutions [1]. The alliance of these two domains conceives the necessary balance and compels investigation regarding the feasibility, trends and scope of using soft computing techniques for cyberbullying detection on the social multimedia data. Recent trends demonstrating the use of soft computing techniques to automate cyberbullying detection has been observed. This makes it necessary to document the preliminary work and review the ongoing work within this domain. The next section describes the process of review adopted for this systematic study.

3 Review process

A systematic literature review (SLR) intends to identify, critically assess and integrate the findings of all pertinent, high quality primary studies addressing specific research questions pertaining to the research domain. This SLR was planned and conducted based on the format put forward by Ketchenham and Charters in 2007 [28]. The overall review process was categorized into six phases as per the guidelines given. The following Fig. 10 depicts the process of SLR.

The first phase ascertains and formulates the research questions (RQ's) within the domain identified for the survey. The next phase, that is, the '*Search strategy*' is designed to recognize and locate the relevant research studies addressing the defined research questions. In the next phase, *'Study selection',* the scope of the study is constricted by using a selection criterion known as '*Inclusion-Exclusion*' criteria. This phase yields the number of the relevant studies that can be included within the specified domain. Next phase is called as '*Quality assessment*' that evaluates the worthiness of the selected studies. The goal is to ensure the quality and similarity of included studies, and clearly define the boundaries of the review. Post this phase is the *'Data extraction'* phase which extracts the relevant and required data in order to answer the specified research questions. It produces a summarized critique to evaluate, extend and/or identify gaps and incosistencies, if any and provide directions for future research. The last *'Data synthesis'* phase summarizes the results of these qualitative filtered studies using

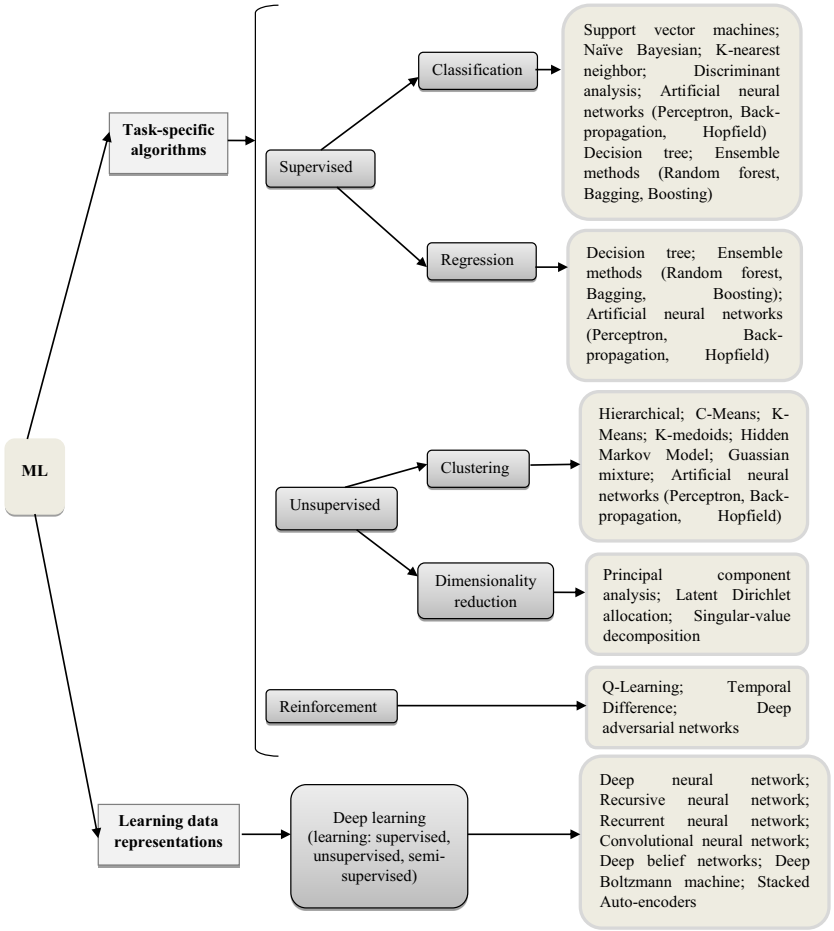


Fig. 7 Machine learning techniques

visualization tools such as graphs, charts and meta-summary tablesThe following sub-sections identify the relevant RQ's which this SLR intends to answer followed by the details of selection and examination of the selected relevant studies.

3.1 Research questions

The following RQs were identified:

- RQ1: Which are the most distinguished and relevant conferences and journals with published studies?
- RQ2: On which datasets and domains the studies using soft computing techniques for cyberbullying detection on social multimedia have been conducted?

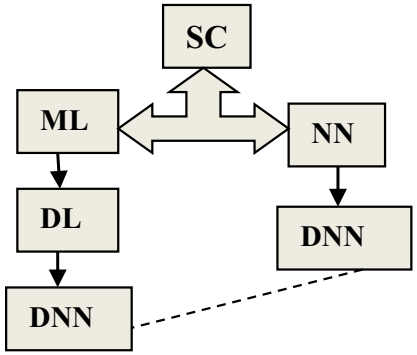


Fig. 8 Relation between SC, ML and DL

- RQ3: Which is the most frequently used soft computing techniques used for cyberbullying detection on social multimedia?
- RQ4: What are the widely applied key performance indicators that are used for evaluating the applied techniques?

3.2 Search strategy

An adept strategy of extensive search (starting the first reported study in April 2003-till date), was set up of in order to extract as many as potential relevant research studies which expound the use of soft computing techniques in cyberbullying detection. This phase fragmented the selected RQ's into individual concepts in order to create the '*search terms'* which could further be searched in the identified databases/e-portals/digital libraries/digital portals. The search terms were identified for this SLR were: cyberbullying, soft computing techniques, machine learning, supervised, unsupervised and social multimedia and were explored in titles, keywords and abstracts of studies to extract all related primary research studies from journals of high repute and highest relevance to the topic of study, available within six prominent digital

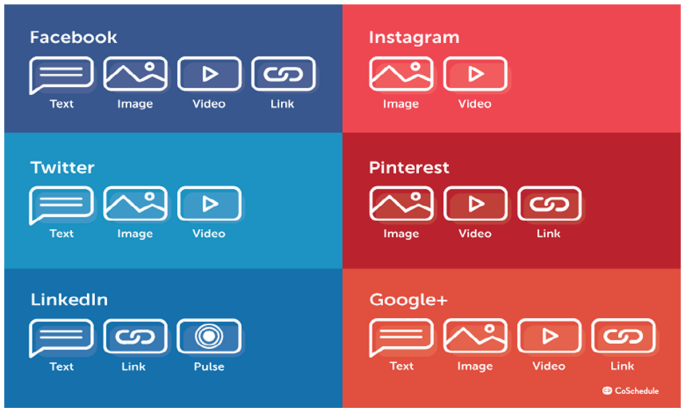


Fig. 9 Multimedia support by popular social networking sites [27]



Fig. 10 Process of SLR

libraries (publishers), namely, ACM, IEEE, Elsevier, Wiley, Springer and Taylor and Francis. The grammatical variations of these terms such as synonyms etc. were also used to search exhaustively. Boolean expressions (or/and) were further used for expanding or narrowing the sweep of the search to filter potentially relevant papers. The reference section of the relevant studies was also examined to extract cross-citations. Thus, the purpose of this step was to identify, select and extract the relevant set of research papers for conducting review.

3.3 Study selection

This phase involved isolating the irrelevant, redundant and non-pertinent studies based on the 'Exclusion-Inclusion' selection criteria. It performed the filtering procedure by either selecting or rejecting the studies which facilitate or directly answer at least one RQ within the selected problem domain. The following Inclusion-Exclusion Criteria was adopted:

Inclusion criteria:

- Studies published in the last 15 years i.e. from April 2003-till date (Sept'2018).
- Studies representative of cyberbullying detection on social multimedia.
- Studies focusing on the application of unsupervised and supervised learning algorithms.
- Studies focusing on the application of supervised machine learning (ML) algorithms like Decision Tree (DT), Support Vector Machine (SVM), k Nearest Neighbor (kNN), Random Forests (RF), Linear Regression (LR), Logistic Regression (LogR), Boosting (Bos), Bagging (Bgg), Adaboost (Adb), Multiple Regression (MR), Maximum Entropy (MaxE) etc. for detecting cyberbullying on social multimedia.
- Studies with unsupervised machine learning algorithms in soft computing such as K-Means Clustering (KMC), C-Means Clustering (CMC), Hierarchical Agglomerative Clustering (HAC) etc.
- Studies including soft computing techniques such as Probabilistic Reasoning which includes Naïve Bayesian (NB) or Bayes Network (BN), Neural Networks (NN), Fuzzy logic (FL), Evolutionary Computing (EC) for cyberbullying detection on social multimedia.
- Studies representing the application of deep learning (DL) techniques like Convolutional Neural Network (CNN) etc. for cyberbullying detection on social multimedia.
- Studies with hybrids of soft computing techniques for detecting cyberbullying on social multimedia.
- Studies involving the comparison of above mentioned techniques.
- Studies involving detecting cyberbullying on social media in English language only.

- Studies involving cyberbullying detection in multimedia like images, texts, videos etc.

Exclusion criteria:

- Studies which are without proper empirical analysis or benchmark comparisons.
- Studies that are purely reviews or surveys or theoretical concepts on cyberbullying detection without any implementations.
- Studies on languages other than English (for example Dutch, Portuguese, Latin, Chinese, Arab, Spanish etc.) and multilingual cyberbullying detection (such as mash-up languages i.e. mixed usage of different languages, for example, Hinglish is a mixture of English and Indian Hindi language).

3.4 Quality assessment

In order to maintain the quality standard of the selected studies a careful consideration had been affirmed by taking the novelty of technique proposed and the technical content (data set and evaluation methods used). The quality check had been imposed in order to evaluate the worthiness, significance and strength of the selected studies based on various w eighing parameters, as discussed next:

- ***Novelty***: to judge whether the proposed technique is a novel one or just an enhancement or improvement over an existing one
- ***Technically content***: to discover the real and clear motivation behind the proposed technique. Also to find whether the scope and limitation of the proposed technique is evident and unambiguous.
- ***Result and analysis***: to assess whether the proposed technique is tested on a standard benchmark data set or a random data set,with proper evaluation of efficacy measures and compared with existing techniques.
- ***Publication***: to identify whether the selected study belongs to a conference or a high impact journal and the number of the citations that the study has. Although not much weightage has been given to this parameter as a recent study may not have many citations.

Thus, each selected study was evaluated out of 10 and scored on the following basis: 2 for novelty, 1.5 for publisher, 5 for results and analysis in which 2 was for data set, 2 for evaluation criteria used and 1 for the comparison with any of the existing technique and the rest 1.5 for technical writing. This qualitative assessment is given in the following Table 2:

3.5 Data extraction

In this phase, finally the key information was extracted from the selected studies and was summarized based on the mapping of the selected study to one or more RQ's. The information acquired from the extracted research studies was: details like authors, year of publication, datasets used, techniques applied, domains targeted, social media which was used for analysis, textual or multimedia content on which the techniques were implemented, type of cross validation that was used, key performance indicators that were employed for evaluation of the techniques, followed by remarks. All this information was then stored in a table for data synthesis.

Table 2 Quality assessment

Quality level	Number of studies	Percentage
Outstanding (9.5 < score < =1)	4	7.69
Excellent (8 < score < = 9.5)	9	17.31
Good (7 < score < =8)	10	19.23
Average (5.5 < score < =7)	18	34.61
BelowAverage(4 < score < =5.5)	6	11.54
Poor (score < =4)	5	9.62

3.6 Data synthesis

The goal of the '*Data Synthesis*' phase is to summarize and interpret the extracted information in order to finally output the result of the SLR as direct answers to the identified RQs using critical analysis, discussions and different visual representations such as tables, graphs, charts, etc. To filter the most significant and qualitative work, search and study selection procedures were carried twice. Meta-analysis then quantitatively combined results of the studies in the SLR. It collated data to generate statistically significant resultsand summaries from the pooled set of relevant studies. Identified search terms were input as search query inthe six selected digital libraries which resulted in 320 papers. After removing redundant studies, 255 studies were obtained on which the inclusion and exclusion criteria was applied. Fifty-two potentially relevant studies were then filtered for further qualtitative analysis out of which 47 high quality studies eventually formed the basis of this review. Figure 11 depicts the overall search process applied in order to fetch the most relevant studies.

4 Literature survey

The review of the final set of studies identified for this SLR which demonstrate the use of soft computing techniques for cyberbullying detection on social multimedia is given in Table 3. As

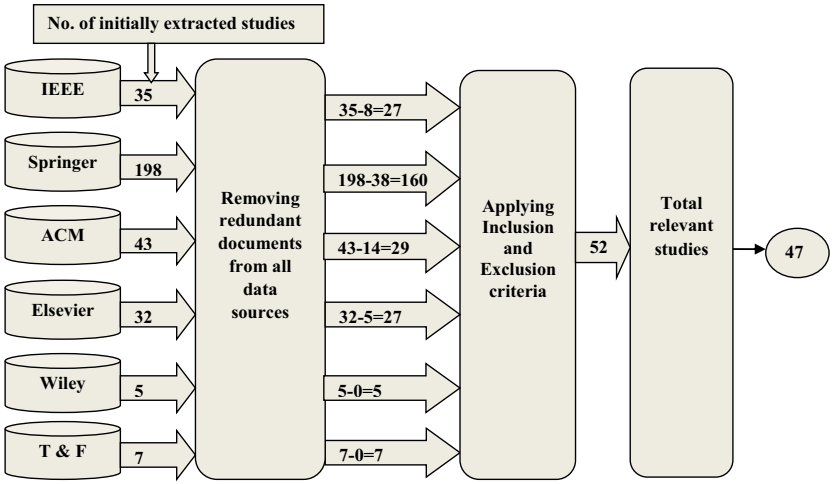


Fig. 11 Systematic literature review procedure

Table 3 Summary details of the studies selected for this review

S.No.	Author	Publication	Year	Techniques	Social Media	Textual or multimedia	Data set
1.	Reynolds et al. [43]	International Conference on Machine Learning and Applications, IEEE	2011	DT, kNN, SVM	Formspring.me	Texts	Author collected data from the Formspring.me and extracted the information from the sites of 18,554 users, containing 2696 posts for training and 1219 for testing purposes.
2.	Nahar et al. [32]	Asia-Pacific Web Conference. Springer, Berlin, Heidelberg	2012	BoW, PLSA, Bayes method, SVM	Kongregate, Slashdot and MySpace	Texts	Data taken from the workshop on Content Analysis (http://caw2.barcelonamedia.org/).
3.	Xu et al. [52]	Conference of the North American chapter of the association for computational linguistics: Human language technologies, ACM	2012	NB, L-SVM (linear), R-SVM (RBF) and LogR, CRF	Twitter	Texts	Data taken from uniformly sampled 990tweets for manual inspection by five experienced annotators.
4.	Kontostathis et al. [30]	Proceedings of the 5th annual acm web science conference. ACM.	2013	BoW, EDLS, tf-idf	Formspring.me	Texts	Author collected data from the Formspring.me and contained 13,652 posts. The data has been labelled using AMT.
5.	Dadvar et al. [17]	European Conference on Information Retrieval, Springer, Berlin, Heidelberg	2013	SVM	YouTube	Texts	Author collected 4626 comments from 3858 distinct users
6.	Sheeba et al. [48]	International Conference on Computational Intelligence and Computing Research, IEEE.	2013	FL, MaxE, CMC, Fuzzy C means, Fuzzy DT	Twitter, Facebook	Texts, audios	Random
7.	Nahar et al. [33]	In Australasian Database Conference Springer, Cham	2014	Naive Bayes multinomial, Stochastic Gradient Descent, RF, LogR, Fuzzy SVM (FSVM), Kernel-based Fuzzy C-Means (K-FCM) clustering	Myspace, Kongregate, and Slashdot.	Texts	Data provided by Fundacion Barcelona Media for the workshop on content analysis (http://caw2.barcelonamedia.org/).
8.	Parime&Suri [37]	International Conference on Circuit, Power and Computing Technologies [ICCPCT], IEEE	2014	SVM	Myspace	Texts	Myspace
9.	Dadvar et al. [18]	Canadian Conference on Artificial Intelligence, Springer, Cham	2014	NB, DT, SVM, MCES (Expert System)	YouTube	Texts	Author collected 54,050 comments from 3825 distinct users
10.	Michalopoulos et al. [31]	Computers & security, Elsevier.	2014	FL, NB, kNN, MaxE, SVM	Perverted-justice website	Texts	Random
11.	Holt et al. [24]	Journal of Criminal Justice, Elsevier	2014	LogR	–	Texts	Author collected a self-administered questionnaire for 6th to 12th grade students in 14 middle and high schools in the Iredell-Statesville School System (ISS) in North Carolina. 1972 students had completed the survey.
12.	Byrne et al. [9]	Journal of Computer-Mediated Communication, Wiley	2014	LogR	General survey conducted	Texts	C + R Research, a professional research firm in Chicago collected data from parents and children. The data was related to the survey involving questions on cyberbullying.
13.	Rafiq et al. [39]	International Conference on Advances in Social Networks Analysis and Mining, ACM	2015	Snowball sampling method, NB, Adb, DT, RF	Vine	Videos	Author collected 652 K media sessions from Vine that contained information such as user id, profile information, videos posted by a user, post id's etc.
14.	Chavan and Shylaja [14]	Advances in computing, communications and informatics (ICACCI), International Conference, IEEE	2015	SVM, LogR	Kaggle	Texts	Authorcollected 2647 comments

Springer

Multimedia Tools and Applications

Table 3 (continued)

S.No.	Author	Publication	Year	Techniques	Social Media	Textual or multimedia	Data set
15.	Balci and Salah [6]	Computers in Human Behavior, Elsevier	2015	DT, SVM, KMC, Bayes Point Machine (BPM)	Okey (CCSoftOkey Player Abuse (COPA) Database)	Texts	Author gathered 800,000 Okey games along with the player interactions in the chat area over a period of six months.
16.	Nandhini and Sheeba [34]	International Conference on Advanced Computing Technologies and Applications (special issue of Procedia Computer Science), Elsevier	2015	NB, FL, FuzGen (hybrid of Fuzzy Logic and GA)	Formspring.me and MySpace	Texts	–
17.	Balakrishnan V [5]	Computers in Human Behavior, Elsevier	2015	LogR	Facebook	Texts	Author prepared a questionnaire of 393 participants consisting of questions related to cyberbullying involving victims and perpetrators, sexting (i.e. sharing sexually suggestive photos or messages through mobile phones and other mobile media), and personalities (i.e. questions related to their overall self-esteem and family).
18.	Zhang Et al. [55]	International Conference on Machine Learning and Applications, IEEE	2016	PCNN, CNN	Twitter and Formspring.me	Texts	Author collected 1313 messages from Twitter and 13,000 messages were Collected from Formspring.me and labeled by a web service called Amazon Mechanical Turk
19.	Zhao et al. [57]	Proceedings of the 17th international conference on distributed computing and networking. ACM.	2016	Continuous Bag of Words, Semantic-enhanced BoW Model, Embeddings-enhanced Bag-of-Words (EBoW), Latent Dirchilet Allocation (LDA),	Twitter	Texts	Author collected 1762 random tweets as on 6 Aug 2011.

Table 3 (continued)

S.No.	Author	Publication	Year	Techniques	Social Media	Textual or multimedia	Data set
				tf-idf, Latent Semantic Analysis (LSA), word embeddings, SVM			
20.	Hosseinmardi et al. [25]	International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE	2016	Snowball sampling method, backward feature selection approach, LogR, ridge regression classifier	Instagram	Images	Author collected data from around 41 K Instagram user ids using a snowball sampling method via the Instagram API.
21.	Gordeev [20]	International Conference on Speech and Computer, Springer, Cham	2016	RF, CNN-non-static, CNN (POS)	Imageboards (4chan.org, 2ch.hk)	Texts	Author collected 1000 English messages
22.	Hammer [23]	International Conference on Industrial Networks and Intelligent SystemsSpringer, Cham.	2016	Logistic LASSO regression	YouTube	Texts	Author collected a total of 24,840 sentences where 1469 sentences were violent
23.	Gordeev [21]	Procedia-Social and Behavioral Sciences, Elsevier	2016	word2vec, NN, RF	Imageboards (4chan.org, 2ch.hk)	Texts	Author collected 1000 English messages
24.	Al-garadi et al. [3]	Computers in Human Behavior, Elsevier	2016	synthetic minority oversampling technique (SMOTE), NB, SVM RF, and kNN	Twitter	Texts	Author collected data between January 2015 and February 2015 and contain 2.5 million geo-tagged tweets. Author randomly selected 10,606 tweets from collected data.
25.	Potha et al. [38]	Knowledge-Based Systems, Elsevier	2016	HAC, Bayesian hierarchical clustering, SVM	Perverted-Justice (PJ)	Texts	–
26.	Rafiq et al. [40]	Social Network Analysis and Mining, Springer	2016	LDA, Adb, DT, RF, Extra tree classifier, SVM (SVM Linear, SVM Polynomial, SVM rbf (radial basis function), SVM Sigmoid), kNN, NB, MLP, LogR	Vine	Videos	Author collected Vine information from 59,560 users about 652 K media sessions.
27.	Papegnies et al. [36]	International Conference on Statistical Language and	2017	Bag of words, TF-IDF, Probability of n gram emission, Context	MMO (online game chat systems)	Texts	Author accessed database containing 4, 029, 343 messages

Table 3 (continued)

S.No.	Author	Publication	Year	Techniques	Social Media	Textual or multimedia	Data set
		Speech Processing. Springer, Cham		based (SVM) and graph based classifier			where 779 messages were flagged by one or more users as abusive and 1558 as non-abusive. Total 2000 random messages were fetched.
28.	Sedano et al. [46]	International Conference on Artificial Intelligence and Soft Computing. Springer, Cham	2017	SVM, FL	Twitter	Texts	Author collected 18,504 tweets from June to December 2016
29.	Thu and New [50]	International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), IEEE	2017	SVM	Twitter, Amazon	Texts	Author collected 674 real news-wire articles and 226 satire newswire articles. Also, they had gathered and pre-processed around 20 K-30 K tweets for each satire and non-satire corpus. Publicly available dataset was also used (https://rpubs.com/pm0kjp/satire_serious_reviews).
30.	Zhao and Mao [56]	IEEE transactions on affective computing	2017	Semantic-enhanced Marginalized Stacked DenoisingAutoencoder - smSDA (deep learning method), mSDA, SVM, LSA, LDA, BoW, sBo W, BWM (Bullying word matching), word embeddings	Twitter and MySpace	Texts	Author collected 7321 tweets from Twitter from 6 Aug 2011 to 31 Aug 2011 and 1539 data samples from MySpace.
31.	Raisi and Huang [41]	International Conference on Advances in Social Networks Analysis and Mining, ACM	2017	Participant-Vocabulary Consistency (PVC) using Alternating Least Squares, snowball sampling	Ask.fm, Instagram, and Twitter	Texts	Author collected 296,308 tweets from Twitter from 1 Nov 2015 to 14 Dec 2015. Author had also gathered 2,863,801 question-answer pairs from Ask.fm and 9,828,760 messages

Table 3 (continued)

S.No.	Author	Publication	Year	Techniques	Social Media	Textual or multimedia	Data set
							from Instagram. For labelling the data, a web service called Amazon Mechanical Turk (AMT) was used.
32.	Chatzakou et al. [11]	Conference on Hypertext and Social Media, ACM	2017	K means, Expectation-maximization algorithm, RF	Twitter	Texts	Author collected 650 K tweets about the Gamergate (GG) controversy. Author had also gathered a baseline dataset with 1 M random tweets.
33.	Chatzakou et al. [12]	International Conference on World Wide Web Companion, ACM	2017	RF	Twitter	Texts	Author collected 650 K tweets on hate related topics and around 1 M baseline tweets random from Jun to Aug 2016.
34.	chatzakou et al. [13]	Proceedings of the 2017 ACM on Web Science Conference. ACM	2017	NB, DT, RF, NN, CBoW	Twitter	Texts	Author collected 650 K tweets on hate related topics and around 1 M baseline tweets random from Jun to Aug 2016.
35.	Garc'ıa--Recuero [19]	International Conference on Advances in Social Networks Analysis and Mining, ACM	2017	MinHashes, DT, RF, Extra Trees, Gradient Boosting, Adb and SVM, Voting	Twitter	Text	Author collected data from 163 trusted humans that provided 14,193 annotations.
36.	Ashktorab et al. [4]	Web Science Conference, ACM	2017	Snowball sampling, Latent Dirichlet Allocation topic modelling, NB	Ask.fm	Texts	Author searched for the key terms that are generally associated with cyberbullying. From Ask.fm and had then fetched random user profile information from it.
37.	Bourgonje et al. [7]	International Conference of the German Society for Computational Linguistics and Language Technology, Springer, Cham	2017	Bayes, Bayes expectation maximization, DT, Multivariate LogR, MaxE, Winnow2, BoW	Twitter, Wikipedia Talk pages	Texts	Author gathered data from Twitter containing 15,979 tweets and from Wikipedia Talk pages containing 11,304 annotated Comments.
38.	Haidar et al. [22]	Cyber Security in Networking Conference IEEE	2017	NB, SVM	Twitter	Texts	Author collected 91,431 tweets

Table 3 (continued)

S.No.	Author	Publication	Year	Techniques	Social Media	Textual or multimedia	Data set
39.	Wint et al. [51]	Digital Information Management (ICDIM), Twelfth International Conference, IEEE	2017	CNN, NB	Twitter, Formspring.me	Texts	Author collected 1,578,627 tweets from Twitter and 18,554 users from Formspring.me
40.	Sarna& Bhatia [45]	International Journal of Machine Learning and Cybernetics,Springer	2017	NB, kNN, DT, SVM	Twitter	Texts	Author collected random tweets via customized crawler written in Python.
41.	Rakib and Soon [42]	Asian Conference on Intelligent Information and Database Systems, Springer, Cham	2018	RF	Reddit, kaggle	Texts	Author collected 6594 raw comments.
42.	Agrawal and Awekar [2]	European Conference on Information Retrieval,ECIR, Springer, Cham	2018	CNN, LSTM, BLSTM, BLSTM with attention, logistic regression, SVM, RF, NB	Formspring, Twitter, Wikipedia	Texts	Author collected 12 k posts for FormSpring, 16 k for Twitter and 100 k for Wikipedia
43.	Chen et al. [15]	Neural Computing and Applications, Springer	2018	SVM, Logistic regression, LSTM+2D TF-IDF, CNN +2D TF-IDF, LSTM+ EMBEDDING, CNN + EMBEDDING	Twitter	Texts	Author collected random aggressive comments from Twitter.
44.	Koban et al. [29]	Computers in Human Behavior, Elsevier	2018	MR	Facebook	Texts	Author conducted an online survey of 256 participants concerning personality dispositions, participants' Internet and Facebook usage, as well as single-item measures of their interest and their level of expertise in four different news subjects (i.e., politics, sport, social issues, and terrorism).
45.	Coletto et al. [16]	World Wide Web Conference 2018,ACM	2018	LogR	Twitter	Texts	Author used twitter datasets. The first one contained 977 k English tweets from Dec 2008 to Jan 2009 and the other contained 1 M English tweets collected in Dec 2015.
46.	Sharma et al. [47]	International Conference on Advances in Computing and Communication Engineering (ICACCE) 2018, IEEE	2018	LogR, SVM, RF, Gradient Boost	Kaggle	Texts	Author collected data that contained2235 samples.
47.	Bu et al. [8]	International Conference on Hybrid Artificial Intelligence Systems, Springer	2018	CNN, LSTM	Kaggle	Texts	Author collected data that contained 8815 comments. Amongst all, 2818 comments were labeled as cyberbullying.

Springer

Multimedia Tools and Applications

S.No.	Domain	Tools	CV	KPI	Remarks
1.	Random	Weka	10	A	DT outperformed the other classifier with A of around 79%.
2.	Games	LibSVM	10	A	Feature selection had shown improved accuracy. SVM performed best with 500 features.
3.	Random	WEKA	5	A, P, R, F	Key problems have been identified in using social multimedia data and formulate them as NLP tasks, including text classification, role labelling, sentiment analysis, and topic modelling.
4.	Random	R	2	P, R, TP	Author proposed a model that resulted in better results for CD.
5.	Movies	–	10	P, R, F	Author observed that incorporation of context in the form of User's activity history improves CD accuracy.
6.	Meeting transcripts	–	–	A	Author obtained improved CD results using FL techniques.
7.	Random	Weka	10	P, R, F	Author proposed a semi-supervised method that had shown improved results as compared to traditional methods for CD.

Table 3 (continued)

S.No.	Domain	Tools	CV	KPI	Remarks
8.	Random	RapidMiner,	10	Cn	Author had taken into account the psychological factors related to cyberbullying for identifying the absence and presence of abusive content.
9.	Random	WEKA 3	10	AUC	Author found that Naive Bayes outperformed the other two algorithms.
10.	Romantic movies, chats	Rainbow (front-to-end document classification tool that performs instant classification of a given text)	10	A, FN, FP	Author developed a 'Grooming Attack Recognition System' for real-time identification, assessment and control of cyber grooming attacks in favor of child protection.
11.	Adolescent problem behavior	–	–	M, SD, Rg, Cr	Author discussed the implications of various demographic factors for policy responses to bullying victimization.
12.	Adolescents problems, pornography or sexual imagery	–	–	A	Author obtained accuracy of more than 70% for the cases where kids are cyberbullied by others.
13.	Well known celebrities	CrowdFlower, NLTK toolkit	10	A, P, R	Amongst all, Adb had obtained the highest accuracy of around 76%.
14.	Random	–	–	P, R, AUC,ACC	Author proposed that the suggested hypothesis increase the accuracy by 4%.
15.	Player demographics, statistics, game records, interactions and complaints	Infer.NET library	10	TP, TN, FP, FN, P, Sn, Sp.	Author proposed a model for assessing different types of features for detecting abuse automatically.
16.	–	–	–	P, A, R, F	Author proposed a hybrid approach for CD where GA is used for optimizing the parameters and to obtain precise output and FL has been used to retrieve relevant data for classification from the input.
17.	Sexting, self-esteem, family.	–	–	MW, KW	Author claimed that the proposed model for cyberbullying was significant where age and gender were found to be insignificant predictors for Cyber-victims and cyberbullies.
18.	Random	eSpeak, Theano package, Python	5 and 10	A, R, P, F, TP, FP, TN, FN	Author proposed a novel PCNN model for CD and the results show that the novel approach had outperformed the existing methods in terms of accuracy.

Table 3 (continued)

S.No.	Domain	Tools	CV	KPI	Remarks
19.	Random	Word2vec	5	P, R, F	Author proposed a novel learning method called as EBoW for CD that yielded enhanced results.
20.	Elite users (famous personalities, like actors, singers etc	–	5	P, R, F, FP, ROC, AUC	Author achieved high performance in predicting cyberbullying using the proposed approach.
21.	Random	–	10	F	Author observed that Random Forest classifier surpassed CNN for the task of detecting aggression for the English language
22.	Religious and political	R (Glmnet package)	–	MSE	Author proposed a method which can automatically detect threats of violence using machine learning.
23.	Random	NLTK-toolkit	10	A	Author proposed a method that detected automatic aggression with 88% accuracy.
24.	Random	WEKA	10	P, R, AUC, F	The results exhibited that the Random forest using SMOTE alone showed the best AUC (0.943) and f-measure (0.936).
25.	Sexual conversations	Matlab	–	R	Author proposed clustering based method for extracting patterns in sexual cyberbullying data and had shown improved results.
26.	Public or user profiles	CrowdFlower	10	A, P, R	RF yielded best A, P and R of more than 85%
27.	Games	Python-iGraph, Sklearn (SVC (C-Support Vector Classification) toolkit	10	P, R, F	Author presented an approach based on graph features for automatically detecting online abuse.
28.	School students and staff members	Java Swing, Twitter streaming API, Tweepy, Mysql database, PHP	–	A	Author presented a model where the output of SVM is fed as input to Fuzzy Logic for identifying the bullying severity.
29.	News-articles and Amazon Products Review s	Tweepy (Python library), SÉANCE (Sentiment analysis and social cognition engine) text analysis tool, Bag of words, term frequency-inverse document frequency (TFIDF), term	10	P, R, F, A	Author proposed an approach for detecting satirical languages in both short text (tweets) and long text (newswire articles and product reviews). Author had shown that the model with supervised weighting TFRF worked better in long text whereas the model with unsupervised weighting TFIDF worked better for short text.

Table 3 (continued)

S.No.	Domain	Tools	CV	KPI	Remarks
		frequency relevance frequency (TFRF), SenticNet			
30.	Random	word2vec	10	A, F	Author proposed an approach called as smSDA for text based CD that showed improved results.
31.	Random	scipy.sparse	3	P	Author proposed a weakly supervised PVC model for CD that had shown enhanced results.
32.	Gamergate Controversy [hate speech] and random	SentiStrength tool	–	P, R, ROC	Author proposed an unsupervised machine learning analysis for better understanding the behaviors of abusive users on social multimedia alike Twitter.
33.	Random and hate related (Gamergate controversy) [hate speech]	–	10	P, R, CK, ROC	The study depicted that the author's approach had produced promising results with high accuracy for detecting aggressive and bully users on Twitter.
34.	Random and hate related (Gamergate controversy) [hate speech]	CrowdFlower, SentiStrength Tool, WEKA	10	P, R, ROC, CK, RMSE	Author proposed a robust approach for understanding the properties of bullies and aggressors on Twitter and improved results were achieved.
35.	Elite users	Crowdflower, Trollslayer	5	P, R, F	MinHashes obtained better abuse detection rates with supervised learning and also minimized the amount of computation as well.
36.	Random	–	–	CK, P, R, F	The proposed approach encouraged the performance of the classifier reasonably for accurate automatic detection of different discourse categories.
37.	Random	McCallum 2002	10	A, P, R, F	Author observed that logistic regression implementation, using word unigrams, outperformed the best scoring feature set in Twitter dataset
38.	Random	WEKA, PHP, python, mongoDB, SentiStrength	–	P, R, F, TP, FP	Author presented a solution for detecting and stopping cyberbullying using SVM and Naive Bayes
39.	Random	–	–	A	Author measured accuracy on collected datasets using CNN and Naïve Bayes.
40.	Random	MATLAB, Python, Part-Of Speech (POS)	–	P, R, F	Author had shown that less users are involved in indirect cyberbullying than direct cyberbullying.
41.	Random	MongoDB, word2vec,Python, GloVe	–	AUC, P	Author depicted that the presented model had2% improvement of precision over the next best score.
42.	Random	GloVe, SSWE	5	P, R, F	

Table 3 (continued)

S.No.	Domain	Tools	CV	KPI	Remarks
					This study analyzedcyberbullying detection on various topics across multiple SMPs using deep learning based models and transfer learning.
43.	Random	Word2Vec, Glove	–	A, Micro-AUC, Macro-AUC	Author achieved improvement in convolutional neural networks (CNN) using 2-dimensional tf-idffeatures.
44.	Personality, politics, sport, social issues, and Terrorism.	–	–	M, SD, Cr	Author presented a study that examined participants' intention to comment in an uncivil manner that typically hinders a productive public discussion.
45.	Random	Face++	–	P, R	The results showed that aggressive users smile less. Also, they appeared not happy in their profile pictures,
46.	Random	Python	–	A, P, R AUC	Results depicted that Logistic Regression and Random Forest performed better than SVM andGradient Boosting.
47.	Random	–	10	ROC, AUC,	Author proposed hybrid architecture of character-level CNN and word-level LSTMthat outperformed other machine learning methods.

discussed in the data extraction phase, the information extracted from the selected studies included details about the authors, publication, year of publication, techniques applied, social media which was selected, type of multimedia on which SC techniques were applied, domains targeted, tools which were used, type of cross validation (CV) used, key performance indicator (KPI) [Accuracy (A), Precision (P), Recall (R), F score/ F1 score/ F1 measure/ F measure (F), Confidence (Cf), Sensitivity (Sn), Specificity (Sp), True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN), ROC, AUC, Cohen's Kappa measure (CK), Root Mean Squared Error (RMSE), MSE, ACC, Mean (M), Standard Deviation (SD), Range (Re), Correlation (Cr), Mann–Whitney U-test (MW), Kruskal–Wallis test (KW)] and the remarks if any.

From these studies, it was observed that instances of cyberbullying can be easily seen and are more flared-up with the widespread use of social media. This digital abuse risks the mental health and social well-being of the netizens, especially victims. Effective method to analyze online interactions between people to detect cyberbullying activities is an important area of research among social media researchers. Various tools and techniques have been identified and assessed for cyberbullying detection and prediction with linguistic, visual, profile, demographic, and activity features. One such consortium of techniques is soft computing (SC) that envisages group of data-driven approaches that have been used in the past decade for cyberbullying detection (CD) on social multimedia (SM).

These studies show that various SC techniques have been used to detect and predict cyberbullying online. Few initial efforts of using supervised machine learning (ML) for CD in Web 2.0 textual content were reported in 2009 [54] and 2011 [43]. Post that, a range of supervised ML techniques such as SVM, NB, DT, RF, Essential Dimensions of Latent Semantic Indexing, Stochastic Gradient Descent, Bayes Point Machine, Ridge Regression, Logistic LASSO Regression, Extra Tree, Gradient Boosting have been used on a variety of SM namely Twitter, Facebook, MySpace, Formspring.me, Kaggle, Kongregate, Slashdot for CD particularly for textual and audio messages. These proposed methods yield enhanced results as compared to traditional methods (lexicon, rule-based methods) for CD. Likewise, an increasing use of SC techniques such as FL, GA, MaxE, NN etc. and Ensemble methods including Adb, Bos, Bgg have been demonstrated for CD on SM for all sorts of multimedia including images, videos, audios and texts. The implementation was done on the data collected from varied social media such as Vine, Instagram, Amazon and YouTube. The work of the authors depicted improved performance in terms of accuracy for these approaches over statistical methods. Few of the studies also presented the use of distance based algorithms such as kNN and regression based methods namely LogR, LR and MR for CD on SM. Some of the studies also focused on the use of unsupervised and semi-supervised SC techniques for CD on SM. These techniques included C means clustering, Probabilistic Latent Semantic Analysis, Conditional Random Fields, Expectation-Maximization, MinHashes, Graph based classifier and Hierarchical agglomerative clustering. Many reported studies also applied the hybrid techniques namely Fuzzy C means (FCM), Fuzzy Decision Tree (FDT), Fuzzy CD detection on SM in order to boost the performance of the system in terms of accuracy etc. More recently, studies on CD on SM utilizing deep learning (DL) techniques such as Convolution Neural Network (CNN) and pronunciation based convolutionneural network (PCNN), marginalized Stacked Denoising Autoencoders (mSDA), Semantic-Enhanced Marginalized Denoising Auto-Encoder (smSDA), Long short-term memory (LSTM) and Bidirectional Long Short Term Memory (BLSTM) have been reported for multimedia messages collected from social media like Twitter and Formspring.me.

5 Results and discussion

In this section, the results obtained as answers to the RQs defined in the SLR are discussed. Table 4 depicts the mapping of the research articles to the respective RQs.

- *Most distinguished and relevant journals/conferences with published studies in the identified research domain (RQ1)*

Reviewing the literature exhaustively helped us identify the journals/conferences within the selected digital libraries and the ones prominently publishing research in the defined problem domain. The following Table 5 demonstrates the distribution of the research articles in the identified journals/conferences, their proportions and corresponding cumulative proportions of the studies included in this SLR.

The following Table 6 suggests the most relevant and distinguished journals/conferences in this domain of study. Amongst the final selected articles for this SLR, majority of them belonged to Elsevier (Computers in Human Behavior, Elsevier), followed by ACM (Web science conference, International Conference on Advances in Social Networks Analysis and Mining), followed by Springer (European Conference on Information Retrieval), IEEE (International Conference on Machine Learning and Applications) and ACM (International Conference on World Wide Web, ACM).

- *Widely used datasets and domains in which the studies for soft computing techniques in cyberbullying detection on social multimedia have been conducted (RQ2)*

Reviewing the pertinent literature it was observed that the research studies have considered a gamut of data sources including publically available datasets. The probed datasets were either publically available, a set of random data fetched in real-time from various SM like Twitter, Vine, Facebook, MySpace etc., or a random set of collected questionnaires/surveys within a selective topic/ subject/ domain. Publically available datasets were also used in the reported studies. One of the data was provided by 'Fundacion Barcelona Media (FBM)' (http://caw2.barcelonamedia.org/) for the workshop on content analysis. FBM supplied five datasets. Amongst all, the author in [32, 33] had used only three of them, primarily belonging to Kongregate, Slashdot and MySpace. The latter was the data available at (https://rpubs.com/pm0kjp/satire_serious_reviews) that comprised of serious and satire reviews from Amazon and Twitter about the news articles and Amazon's products reviews [50], gathered by Payton and Weigandt for their project 'Satire and Data Science: An Exploration into one of the Current Final Frontiers'.

Table 4 Mapping of RQ's with the relevant research articles

#RQ addressed	Research reference
RQ1	[3, 10, 14, 19, 20, 27, 32, 33, 38, 40]
RQ2	[2–9, 11–25, 29–34, 36–43, 45–48, 50–52, 55–57] (http://caw2.barcelonamedia.org/, https://rpubs.com/pm0kjp/satire_serious_reviews)
RQ3	[2, 3, 6, 14, 15, 17–19, 22, 31, 32, 36–38, 40, 45–47, 50, 52, 56, 57] (https://rpubs.com/pm0kjp/satire_serious_reviews)
RQ4	[2–4, 6, 7, 11–14, 16, 17, 19, 22, 25, 30, 33, 34, 38, 39, 41, 42, 45, 47, 50, 52, 55, 57]

Table 5 Distribution of the research articles in regard to the respective journals/conferences

S. No.	Journal/Conference name	Description	#papers	Proportion (%)	Cumulative Proportion (%)
1.	Elsevier Journals	• Computers & security, Elsevier.	1	17	17
		• Journal of Criminal Justice, Elsevier	1		
		• Computers in Human Behavior, Elsevier	4		
		• Procedia-Social and Behavioral Sciences, Elsevier	1		
		• Knowledge-Based Systems, Elsevier	1		
2.	Elsevier Conferences	• International Conference on Advanced Computing Technologies and Applications (special issue of Procedia Computer Science), Elsevier	1	2	19
3.	Springer Journals	• Social Network Analysis and Mining, Springer	1	6	25
		• International Journal of Machine Learning and Cybernetics, Springer	1		
		• Neural Computing and Applications, Springer	1		
4.	Springer Conferences	• Asia-Pacific Web Conference. Springer, Berlin, Heidelberg	1	26	51
		• European Conference on Information Retrieval, Springer, Berlin, Heidelberg	2		
		• In Australasian Database Conference Springer, Cham	1		
		• Canadian Conference on Artificial Intelligence, Springer, Cham	1		
		• International Conference on Speech and Computer, Springer, Cham	1		
		• International Conference on Industrial Networks and Intelligent Systems, Springer, Cham.	1		
		• International Conference on Statistical Language and Speech Processing. Springer, Cham	1		
		• International Conference on Artificial Intelligence and Soft Computing. Springer, Cham	1		
		• International Conference of the German Society for Computational Linguistics and Language Technology, Springer, Cham	1		
		• Asian Conference on Intelligent Information and Database Systems, Springer, Cham	1		
		• International Conference on Hybrid Artificial Intelligence Systems, Springer	1		
5.	IEEE Journals	• IEEE Transactions on Affective Computing	1	2	53
6.	IEEE Conferences	• International Conference on Computational Intelligence and Computing Research, IEEE	1	21	74
		• International Conference on Circuit, Power and Computing Technologies [ICCPCT], IEEE	1		
		• Advances in computing, communications and informatics (ICACCI), International Conference, IEEE	1		
		• International Conference on Machine Learning and Applications, IEEE	2		
		• International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE	1		
		• International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), IEEE	1		
		• Cyber Security in Networking Conference IEEE	1		
		• Digital Information Management (ICDIM), Twelfth International Conference, IEEE	1		
		• International Conference on Advances in Computing and Communication Engineering (ICACCE) 2018, IEEE	1		
7.	ACM Conferences	• Conference of the North American chapter of the association for computational linguistics: Human language technologies, ACM	1	24	98
		• Web science conference. ACM.	3		
		• International Conference on Advances in Social Networks Analysis and Mining, ACM	3		
		• Proceedings of the 17th international conference on distributed computing and networking. ACM.	1		
		• Conference on Hypertext and Social Media, ACM	1		
		• International Conference on World Wide Web, ACM	2		
8.	Wiley Journals	•Journal of Computer-Mediated Communication, Wiley	1	2	100

Many reported researches were carried on the data fetched directly from various social media using their respective API's. The fetched corpuses were from a variety of domains, topics and time period (referred as topic specific/topic oriented). These prominently belonged to domains including games and sports [6, 29, 32, 36], entertainment [17, 31], meeting transcripts [48], chats [31], adolescent behaviors and their problems [9, 24], pornography [9], sexual imagery [9], elite personalities and well known celebrities including singers, actors

Table 6 Relevant and distinguished journals/conferences

S. No.	Journal/Conference Name	# papers
1.	• Computers in Human Behavior, Elsevier (Journal)	4
2.	• Web Science Conference, ACM	3
	• International Conference on Advances in Social Networks Analysis and Mining, ACM	3
3.	• European Conference on Information Retrieval, Springer	2
	• International Conference on Machine Learning and Applications, IEEE	2
	•International Conference on World Wide Web, ACM	2

etc. s [19, 25, 29, 39], player demographics [6], statistics [6], complaints [6], sexting or sexual conversations [5, 38], family issues [5], self-esteem behaviors [5], religion [23], politics [23, 29], public profiles [40], user profiles [40], education [46], news [50], consumer products [50], hate speech controversy [11–13], social issues [29] and terrorism [29], etc. The following Fig. 12 depicts the year-wisetrend of published work from various domains taken as dataset for empirical evaluation.

Various other reported studies use random datasets [2–4, 7, 8, 14–16, 18, 20–22, 30, 31, 33, 37, 41, 42, 45, 47, 51, 52, 55–57]. Few studies reported self-administered questionnaires/ surveys conducted by the author involving questions related to cyberbullying comprising of victims and perpetrators [5, 9, 24, 29]. The following Fig. 13 shows the year-wise trend of published work with 'topic specified/ topic oriented' datasets from various domains.

The use of random data sets on general topics from various domains has been considered for research evaluations, especially since the year 2011. The following Fig. 14 specifies the wide array of topics from which the randomized datasets have been taken.

Further, the Table 7 identifies the wide list of SM covered for analyzing studies involving the application of SC techniques for CD.

The following Table 8 enumerates the type of multimedia upon which CD was done using SC techniques.

Likewise, Table 9 provides details about the types of tools and software libraries used for carrying out the analyses and implementation of CD on SM using SC techniques.

- *Most frequently used soft computing techniques for achieving efficient results for cyberbullying detection on social multimedia on SM (RQ3)*

Table 10 illustrates the year wise usage of various SC techniques for CD on SM over a wide range of domains.

Thus, the most frequently used soft computing technique for achieving efficient results for cyberbullying detection on social multimedia is machine learning (support vector machine) followed by deep learning techniques. The graph shown in Fig. 15 depicts the quantitative extent of the use of various soft computing techniques for cyberbullying detection on social media.

Apart from these, other subset of SC techniques that were used for analyzing CD on SM are categorized under the SC consortium as follows: *supervised learning, un-supervised learning and semi-supervised learning*. Techniques like Essential Dimensions of Latent Semantic Indexing (EDLSI) [30], Stochastic Gradient Descent (SGDes) [33], Bayes Point Machine

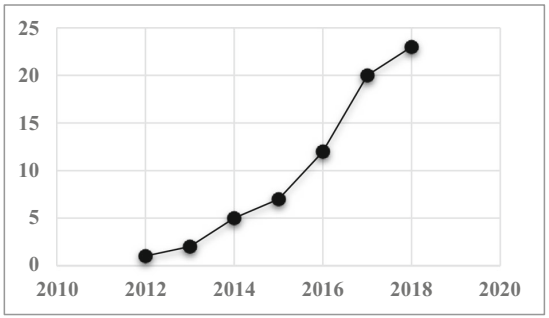


Fig. 12 Year wise cumulative assessment of topic oriented fetched datasets

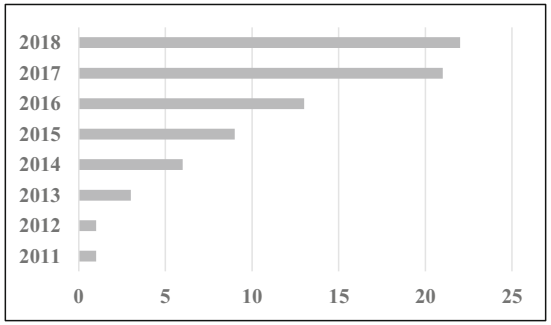


Fig. 13 Year-wise cumulative assessment of random (general) datasets fetched from various SM

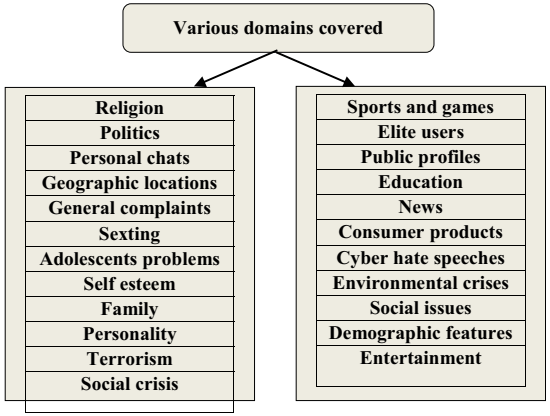


Fig. 14 Various domains covered for research for CD on SM using SC techniques

Table 7 List of social multimedia covered

Social Multimedia	# Papers	Referenced study
Twitter	20	[2, 3, 7, 11–13, 15, 16, 19, 22, 41, 42, 45, 46, 48, 50–52, 55–57] (https://rpubs.com/pm0kjp/satire_serious_reviews)
Forums	9	[2, 6, 7, 20, 21, 32, 33, 36, 50]
Formspring.com	6	[2, 30, 34, 43, 51, 55]
MySpace	5	[32–34, 37, 56]
Kaggle	4	[8, 14, 42, 47]
Facebook	3	[5, 29, 48]
YouTube	3	[17, 18, 23]
Perverted Justice Website	2	[31, 40]
Vine	2	[39, 40]
Ask.fm	2	[4, 52]
Instagram	2	[25, 52]
Slashdot	2	[32, 33]
Reddit	1	[42]

Table 8 List of multimedia covered

Multimedia	# Papers	Referenced Study
Texts	44	[2–9, 11–24, 29–34, 36–38, 41–43, 45–48, 50–52, 55–57] (http://caw2.barcelonamedia.org/, https://rpubs.com/pm0kjp/satire_serious_reviews)
Videos	2	[39, 40]
Audios	1	[48]
Images	1	[25]
Emoticons	0	–
GIF's	0	–

(BPM) [6], Ridge Regression (RR) [25], Logistic LASSO Regression (LLReg) [23], Gradient Boosting (GBos) [19], Multilevel Regression (MR) [29], Extra Tree classifier (ET) [19, 40] and Multilayer Perceptron (MLP) [40] comes under the supervised learning whereas C means clustering (CMC) [48], Probabilistic Latent Semantic Analysis (PLSA) [32], Conditional Random Fields (CRF) [52], Expectation-Maximization (ExpM) [11], MinHashes (MinH) [19], Graph based classifier [36] and Hierarchical agglomerative clustering (HAC) [38] lies under the category of un-supervised learning approaches. Many reported studies also applied the hybrid techniques namely Fuzzy C means (FCM) [33, 48], Fuzzy Decision Tree (FDT) [48], Fuzzy SVM (FSVM) [33], FuzGen (hybrid of Fuzzy Logic and GA) [34] and Bayesian hierarchical clustering (BHC) [38]. Furthermore, few of the studies had also reported the applicability of deep learning (DL) techniques for CD on SM such as pronunciation based convolution neural network (PCNN) [55], marginalized Stacked DenoisingAutoencoders (mSDA) [56], Semantic-Enhanced Marginalized Denoising Auto-Encoder (smSDA) [56],

Table 9 List of tools and software libraries used

Tools	# Papers	Referenced study
Python	8	[22, 36, 42, 45–47, 50, 55]
WEKA	7	[3, 13, 18, 22, 33, 43, 52]
Word2vec	4	[15, 42, 56, 57]
CrowdFlower	4	[13, 19, 39, 40]
SentiStrength	3	[11, 13, 22]
GloVe	2	[2, 15, 42]
MATLAB	2	[38, 45]
NLTK	2	[21, 39]
R	2	[23, 30]
RapidMiner	1	[37]
Rainbow	1	[31]
.NET	1	[6]
LibSVM	1	[32]
ESpeakTheano	1	[55]
IGraph	1	[36]
SKLearn	1	[36]
Java	1	[46]
Séance	1	[50]
Scipy.sparse	1	[41]
Trollslayer	1	[19]
McCallum	1	[7]
SSWE	1	[2]
Face++	1	[16]

Table 10 Year wise distribution of applicability of the SC techniques for CD on SM

Technique	#papers	Year	Referenced study
SVM	27	2012-2018	[2, 3, 6, 14, 15, 17–19, 22, 31, 32, 36–38, 40, 43, 45–47, 50, 52, 56, 57]
DL	15	2016-2018	[2, 8, 15, 20, 51, 55, 56]
NB	14	2012, 2014-2018	[2–4, 13, 18, 22, 31, 33, 34, 39, 40, 45, 51, 52]
RF	13	2014-2018	[2, 3, 12, 13, 19–21, 33, 39, 40, 42, 47]
LogR	13	2012, 2014-2018	[2, 5, 7, 9, 14–16, 24, 25, 33, 40, 47, 52]
DT	9	2014-2017	[6, 7, 13, 18, 19, 39, 40, 43, 45]
kNN	5	2016-2017	[3, 31, 40, 43, 45]
FL	4	2013-2015, 2017	[31, 34, 46, 48]
LDA	4	2016-2017	[4, 40, 56, 57]
MaxE	3	2013-2014, 2017	[7, 31, 48]
Adb	3	2015-2017	[19, 39, 40]
NN	2	2016-2017	[13, 21]
LSA	2	2016-2017	[56, 57]
KMC	2	2015, 2017	[6, 11]
Bys	2	2012, 2017	[7, 32]

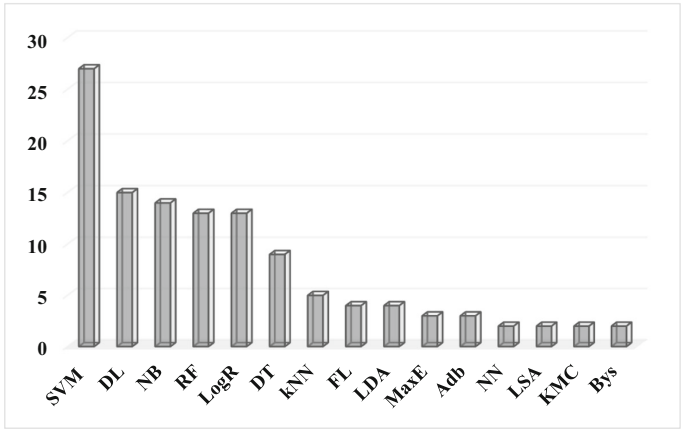


Fig. 15 Quantitative extent of the use of various SC techniques

Long short-term memory (LSTM) [15] andBidirectional Long Short Term Memory (BLSTM) [15]. These hybrid techniques utilizing the combination of supervised and un-supervised methods comes under semi-supervised techniques, such as Bayes Expectation Maximization

Table 11 Year wise distribution of the use of hybrid techniques

S. no.	Technique	#papers	Year	Referenced study
1.	FuzGen (Fuzzy Logic + GA)	1	2015	[34]
2.	FCM (Fuzzy Logic + C Means Clustering)	1	2013-2014	[33, 48]
3.	FDT (Fuzzy Logic + Decision Tree)	1	2013	[48]
4.	FSVM (Fuzzy Logic + Support Vector Machines)	1	2014	[33]
5.	BHC (Bayesian Classification + Hierarchical Clustering)	1	2016	[38]

Table 12 Key performance indicators used

S. no.	KPI	Description	Referenced study
1.	Precision (P)	It defines the exactness of any classifier. A higher precision value indicates fewer 'false positives' (FP) and vice versa. It is given as the ratio of true positives (TP) to all the predicted positives.	[2–4, 6, 7, 11–14, 16, 17, 19, 22, 25, 30, 33, 34, 36, 39–42, 45, 47, 50, 52, 55, 57]
2.	Recall (R)/Sensitivity (Sn)	It defines the sensitivity or the completeness of any classifier. A higher recall value indicates less 'false positives' and vice versa. Recall and precision are bounded by inverse relation with each other. It is given as the ratio of TP to all the actual positives (TP + FN)	[2–4, 7, 11–14, 16, 17, 19, 22, 25, 30, 33, 36, 38–40, 45, 47, 50, 52, 55, 57]
3.	F score/F1 score/F measure/F1 measure/F1 value(F)	It is defined as 'weighted' harmonic mean of Recall and Precision. It is the combination of the precision and recall measures.	[2–4, 7, 17, 19, 20, 22, 25, 33, 34, 36, 45, 50, 52, 55–57]
4.	Accuracy (A)	It is defined as proximity of a measurement to its true value. It is calculated as a proportion of TP and true negatives (TN) among total inspected cases.	[7, 9, 15, 21, 31, 32, 34, 39, 40, 43, 46–48, 50–52, 55, 56]
5.	AUC (Area under the curve)	It is interpreted as AUROC (area under the receiver operating characteristics curve) and is also used for measuring the efficacy of the classifiers based on two metrics i.e. true positive rate and false positive rate.	[3, 8, 14, 18, 25, 42, 47]
6.	ROC	ROC curve plots sensitivity as a function of Specificity for different cut-off points.	[8, 11–13, 25]
7.	Cohen's Kappa (CK)	This metric performs a comparison between the observed and the expected accuracy. It is used for evaluating performance of a single classifier and also among different classifiers.	[4, 12, 13]
8.	Mean (M)	Average of numbers is defined as mean. It is calculated by finding sum of all numbers and then divided the sum by count of numbers.	[24, 29]
9.	Standard Deviation (SD)	It measures the dispersion of data values from mean. Square root of variance gives SD.	[24, 29]
10.	Co-relation (Cr)	It is a measure that shows the degree to which two or more variables vary together.	[24, 29]
11.	Specificity (Sp)	Sp is referred as true negative rate and is denoted as ratio of TN to all the actual negatives (TN + FP).	[6]
12.	Confidence (Cn)	Cnis the probability of prediction of test sample to be in each class.	[37]
13.	Range (Rg)	It is defined as difference between highest and lowest values.	[24]
14.	ACC	It ensures that the label set predicted for a sample must tally with corresponding label set in ground truth labels.	[14]
15.	Mann-Whitney U-Test (MW)	It is used to compare whether two sample means are equal or not.	[5]
16.	Kruskal-Wallis Test (KW)	It is a non-parametric test to determine whether different samples emerge from same distribution.	[5]
17.	MSE	Mean Squared Error is the average of squared differences between actual and predicted values.	[23]
18.	Root Mean Squared Error (RMSE)	It is defined as square root of Mean Squared Error.	[13]
19.	Micro AUC	It is the area under ROC curve.	[15]
20.	Macro AUC	It is the average AUC of the distinct ROC curves for each class	[15]

(BExpM) [7], Fuzzy C means (FCM) [33, 48], Participant-Vocabulary Consistency (PVC) using Alternating Least Squares [41] etc. The following Table 11 presents the studies with hybrid techniques.

- *Widely used performance measures (RQ4)*

The following performance measures were observed in all selected studies to measure the performances of the applied SC techniques. They are referred to as the 'key performance indicators' (KPIs) or 'performance parameters' or 'efficacy measures'. Table 12 illustrates the KPI's used so far.

6 Conclusion

Cyberbullying can take many forms; however, it typically refers to repeated and hostile behavior online to intentionally and repeatedly harass or harm individuals. With the pervasive use of social media, cyberbullying is becoming rampant. Conventional methods to combat cyberbullying included guidelines on cyber-ethics, human moderators, and blacklisting based on the use of profane words. These methods are incapable to deal with the mounting velocity, volume and variety of data generated by the social media, thus necessitating the design and development of novel learning-based computational models. A systematic literature review was conducted on the use of soft computing techniques to detect cyberbullying activities across various social media domains to understand the theory, research and practice trends within the domain. Combining the results of individual studies, as a meta-summary, it was observed that the use of soft computing techniques for cyberbullying detection provided the intelligent analytic paradigm essential for predicting bullying behaviors and activities on both textual and non-textual social media. It is a promising direction of research with practical domain which primarily relies on exploring and understanding the extensibility of human expressions via both textual and non-textual unstructured web-data available. Based on the review, the following research gaps have been identified:

- Detecting and predicting cyberbullying behavior is a non-trivial task with anonymous cyberbullying as an added concern.
- Although researchers are keen in applying soft computing techniques, only few approaches have been explored. Techniques such as deep learning, neural network, ensemble methods, evolutionary computing and hybrids including neuro-fuzzy models have been least explored to substantiate their influence on CD.
- The existing models using SC techniques for CD on SM have majorly considered Twitter, MySpace, Formspring.me, Facebook as the database, making other SM technologies such as Reddit, Vine, Instagram, Flickr, Tumblr, Ask.fm, YouTube etc. open to further application and testing.
- Extracting, selecting and modeling computational features such as linguistic, visual, socio-demographic features (like person's economic status, age, gender, etc.), socio-ecological features (like parental monitoring, hours spent on Internet, racial/community differences etc.) and activity features (behavioral factors) needs further concurrence of soft computing techniques with natural language models and network analysis techniques.
- Most of the reported work done is to detect cyberbullying activity is using textual content, whereas other media types such as audio, video, images are open to research initiatives.

Also, the use of animated GIF's, memes has recently been reported to embarrass or target people on social media, making it an open area of research.

- The informal, short, noisy and unstructured social media further add to the challenges. The use of slangs, mal-formed or colloquial words,mash-up languages (mixed usage of different languages, for example, Hinglish is a mixture of English and Indian Hindi language) make detection of online bullying activities tricky and computationally hard.

Thus, the need to exploit new computational models to detect and predict cyberbullying on social multimedia is abundant making this domain of study a potentially active and dynamic for social media researchers. It compels to look for models that combine the cognition, intelligence and self-tuning behavior of soft computing techniques with disciplines like natural language processing, psychology and artificial intelligence.

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

References

1. Aggarwal CC (2018) Neural networks and deep learning: a textbook. Springer, Berlin
2. Agrawal S, Awekar A (2018) Deep learning for detecting cyberbullying across multiple social media platforms. In: European conference on information retrieval. Springer, Cham, pp 141–153
3. Al-garadi MA, Varathan KD, Ravana SD (2016) Cybercrime detection in online communications: the experimental case of cyberbullying detection in the twitter network. Comput Hum Behav 63:433–443
4. Ashktorab Z, Haber E, Golbeck J, Vitak J (2017) Beyond cyberbullying: self-disclosure, harm and social support on ASKfm. In: Proceedings of the 2017 ACM on Web Science Conference, p 3-12
5. Balakrishnan V (2015) Cyberbullying among young adults in Malaysia: the roles of gender, age and internet frequency. Comput Hum Behav 46:149–157
6. Balci K, Salah AA (2015) Automatic analysis and identification of verbal aggression and abusive behaviors for online social games. Comput Hum Behav 53:517–526
7. Bourgonje P, Moreno-Schneider J, Srivastava A, Rehm G (2017) Automatic classification of abusive language and personal attacks in various forms of online communication. In: International conference of the German Society for Computational Linguistics and Language Technology. Springer, Cham, pp 180–191
8. Bu SJ, Cho SB (2018) A hybrid deep learning system of CNN and LRCN to detect cyberbullying from SNS comments. International Conference on Hybrid Artificial Intelligence Systems 2018, Springer, p 561–572
9. Byrne S, Katz SJ, Lee T, Linz D, McIlrath M (2014) Peers, predators, and porn: predicting parental underestimation of children's risky online experiences. J Comput-Mediat Commun 19(2):215–231
10. Campbell MA (2005) Cyber bullying: an old problem in a new guise? J Psychol Couns Sch 15(1):68–76
11. Chatzakou D, Kourtellis N, Blackburn J, De Cristofaro E, Stringhini G, Vakali A (2017a) Hate is not binary: studying abusive behavior of# gamergate on twitter. In Proceedings of the 28th ACM conference on hypertext and social media, p 65-74
12. Chatzakou D, Kourtellis N, Blackburn J, De Cristofaro E, Stringhini G, Vakali A (2017b) Detecting aggressors and bullies on Twitter. In: Proceedings of the 26th International Conference on World Wide Web Companion, p 767-768
13. Chatzakou D, Kourtellis N, Blackburn J, De Cristofaro E, Stringhini G, Vakali A (2017c) Mean birds: Detecting aggression and bullying on twitter. In: Proceedings of the 2017 ACM on web science conference, p 13-22
14. Chavan VS, Shylaja SS (2015) Machine learning approach for detection of cyber-aggressive comments by peers on social media network. In: Advances in computing, communications and informatics (ICACCI), 2015 International Conference on IEEE, p 2354-2358
15. Chen J, Yan S, Wong KC (2018) Verbal aggression detection on twitter comments: convolutional neural network for short-text sentiment analysis. Neural Comput & Applic:1–10
16. Coletto M, Lucchese C, Orlando S (2018) Do violent people smile: social media analysis of their profile pictures. In: Companion of the Web Conference 2018. International World Wide Web Conferences Steering Committee, ACM, p 1465–1468

17. Dadvar M, Trieschnigg D, Ordelman R, de Jong F (2013) Improving cyberbullying detection with user context. In: European conference on information retrieval. Springer, Berlin, Heidelberg, pp 693–696
18. Dadvar M, Trieschnigg D, de Jong F (2014) Experts and machines against bullies: a hybrid approach to detect cyberbullies. In: Canadian conference on artificial intelligence. Springer, Cham, pp 275–281
19. García-Recuero Á (2017) Efficient privacy-preserving adversarial learning in decentralized online social networks. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ACM, p 1132–1135
20. Gordeev D (2016a) Detecting state of aggression in sentences using CNN. In: International Conference on Speech and Computer, Springer, Cham, p 240-245
21. Gordeev D (2016b) Automatic detection of verbal aggression for Russian and American image boards. Procedia Soc Behav Sci 236:71–75
22. Haidar B, Chamoun M, Serhrouchni A (2017) Multilingual cyberbullying detection system: Detecting cyberbullying in Arabic content. In: Cyber Security in Networking Conference (CSNet), IEEE, p 1-8
23. Hammer HL (2016) Automatic detection of hateful comments in online discussion. In: International Conference on Industrial Networks and Intelligent Systems, Springer, Cham, p 164-173
24. Holt TJ, Turner MG, Exum ML (2014) The impact of self control and neighborhood disorder on bullying victimization. J Crim Just 42(4):347–355
25. Hosseinmardi H, Rafiq RI, Han R, Lv Q, Mishra S (2016) Prediction of cyberbullying incidents in a media-based social network. In: Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, p 186–192
26. http://www.ryanpatrickhalligan.org/ Accessed 14 July 2018
27. https://coschedule.com/ Accessed 14 July 2018
28. Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering. Tech Rep EBSE 1:1–57
29. Koban K, Stein JP, Eckhardt V, Ohler P (2018) Quid pro quo in web 2.0. Connecting personality traits and Facebook usage intensity to uncivil commenting intentions in public online discussions. Comput Hum Behav 79:9–18
30. Kontostathis A, Reynolds K, Garron A, Edwards L (2013) Detecting cyberbullying: query terms and techniques. In: Proceedings of the 5th annual ACM web science conference, p 195-204
31. Michalopoulos D, Mavridis I, Jankovic M (2014) GARS: real-time system for identification, assessment and control of cyber grooming attacks. Computers & Security 42:177–190
32. Nahar V, Unankard S, Li X, Pang C (2012) Sentiment analysis for effective detection of cyber bullying. Asia-Pacific Web Conference, Springer, Berlin, Heidelberg, p 767-774
33. Nahar V, Al-Maskari S, Li X, Pang C (2014) Semi-supervised learning for cyberbullying detection in social networks. In: Australasian database conference. Springer, Cham, pp 160–171
34. Nandhini BS, Sheeba JI (2015) Online social network bullying detection using intelligence techniques. Procedia Computer Science 45:485–492. International Conference on Advanced Computing Technologies and Applications (ICACTA)
35. National Bullying Prevention Center https://www.pacer.org/bullying/. Accessed 26 July 2018
36. Papegnies E, Labatut V, Dufour R, Linares G (2017) Graph-based features for automatic online abuse detection. In: International conference on statistical language and speech processing. Springer, Cham, pp 70–81
37. Parime S, Suri V (2014) Cyberbullying detection and prevention: data mining and psychological perspective. In: Circuit, Power and Computing Technologies (ICCPCT), 2014 International Conference IEEE, p 1541-1547
38. Potha N, Maragoudakis M, Lyras D (2016) A biology-inspired, data mining framework for extracting patterns in sexual cyberbullying data. Knowl-Based Syst 96:134–155
39. Rafiq RI, Hosseinmardi H, Han R, Lv Q, Mishra S, Mattson SA (2015) Careful what you share in six seconds: detecting cyberbullying instances in Vine. In: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ACM, p 617–622
40. Rafiq RI, Hosseinmardi H, Mattson SA, Han R, Lv Q, Mishra S (2016) Analysis and detection of labeled cyberbullying instances in vine, a video-based social network. Soc Netw Anal Min 6(1):88
41. Raisi E, Huang B (2017) Cyberbullying detection with weakly supervised machine learning. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ACM, p 409–416
42. Rakib TB, Soon LK (2018) Using the Reddit Corpus for cyberbully detection. In: Asian conference on intelligent information and database systems. Springer, Cham, pp 180–189
43. Reynolds K, Kontostathis A, Edwards L (2011) Using machine learning to detect cyberbullying. In: Machine learning and applications and workshops (ICMLA), 2011 10th International Conference, IEEE 2:241-244

44. Salawu S, He Y, Lumsden J (2017) Approaches to automated detection of cyberbullying: a survey. IEEE Trans Affect Comput 1:1–20
45. Sarna G, Bhatia MP (2017) Content based approach to find the credibility of user in social networks: an application of cyberbullying. Int J Mach Learn Cybern 8(2):677–689
46. Sedano CR, Ursini EL, Martins PS (2017) A bullying-severity identifier framework based on machine learning and fuzzy logic. In: International conference on artificial intelligence and soft computing. Springer, Cham, pp 315–324
47. Sharma HK, Kshitiz K, Shailendra (2018) NLP and machine learning techniques for detecting insulting comments on social networking platforms. International Conference on Advances in Computing and Communication Engineering (ICACCE) 2018, IEEE, p 265-272
48. Sheeba JI, Vivekanandan K (2013) Low frequency keyword extraction with sentiment classification and cyberbully detection using fuzzy logic technique. In: IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), p 1–5
49. The National Crime Prevention Cyberbullying. http://www.ncpc.org/cyberbullying. Accessed 10 March 2018
50. Thu PP, New N (2017) Implementation of emotional features on satire detection. In Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2017 18th IEEE/ACIS International Conference, IEEE, p 149-154
51. Wint ZZ, Ducros T, Aritsugi M (2017) Spell corrector to social media datasets in message filtering systems. In: Digital Information Management (ICDIM), 2017 Twelfth International Conference, IEEE, p 209–215
52. Xu JM, Jun KS, Zhu X, Bellmore A (2012) Learning from bullying traces in social media. In: Proceedings of the 2012 conference of the north American chapter of the association for computational linguistics: human language technologies, Association for Computational Linguistics, p 656-666
53. Ybarra M (2010) Trends in technology-based sexual and non-sexual aggression over time and linkages to nontechnology aggression. National Summit on Interpersonal Violence and Abuse Across the Lifespan: Forging a Shared Agenda
54. Yin D, Xue Z, Hong L, Davison BD, Kontostathis A, Edwards L (2009) Detection of harassment on web 2.0. Proceedings of the Content Analysis in the WEB 2:1-7
55. Zhang X, Tong J, Vishwamitra N, Whittaker E, Mazer JP, Kowalski R, Hu H, Luo F, Macbeth J, Dillon E (2016) Cyberbullying detection with a pronunciation based convolutional neural network. In: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), p 740–745
56. Zhao R, Mao K (2017) Cyberbullying detection based on semantic-enhanced marginalized denoising auto-encoder. IEEE Trans Affect Comput 8(3):328–339
57. Zhao R, Zhou A, Mao K (2016) Automatic detection of cyberbullying on social networks based on bullying features. In: Proceedings of the 17th international conference on distributed computing and networking, p 43-48



Dr. Akshi Kumar is an Assistant Professor in the Department of Computer Science & Engineering at Delhi Technological University (formerly Delhi College of Engineering). She has been with the university for the past 10 years. She has received her doctorate degree in Computer Engineering in the area of Web Mining, from Faculty of Technology, University of Delhi in 2011. She completed her M. Tech (Master of Technology) with honors in Computer Science & Engineering from Guru Gobind Singh Indraprastha University, Delhi in 2005. She received her BE (Bachelor of Engineering) degree with distinction in Computer Science & Engineering from Maharishi Dayanand University, Rohtak in 2003. Dr. Kumar's research interests are in the area of Intelligent Systems, User-generated Big-data, Social Media Analytics and Soft Computing. She has many publications to her credit in various International Journals with high impact factor and International Conferences with best paper awards. She has authored books & book-chapters within her domain of interest. She is actively involved in research activities as the editorial review board member for international journals, session chair & technical program committee member for international conferences and providing research supervision at under-graduate, post-graduate and doctorate levels. She is a currently a member of IEEE, IEEE (WIE) and ACM and a life member of IACSIT, IAENG, ISTE and CSI.



Mr. Nitin Sachdeva is a Research Scholar in the Department of Computer Science & Engineering at Delhi Technological University, Delhi, India. He completed his Masters in Information Security [MTech (IS), 2014-2016] from the Department of Computer Science & Engineering at Dr. BR Ambedkar National Institute of Technology, Jalandhar, India. He pursued his Bachelors in Information Technology from USIT, Kashmere Gate, Main Campus, Delhi, India. His research interests include Web Technologies, Soft Computing, Intelligent Systems, Text Mining and Social Web Analytics.

ISSN 1064-2269, Journal of Communications Technology and Electronics, 2018, Vol. 63, No. 12, pp. 1424–1430. © Pleiades Publishing, Inc., 2018.

THEORY
OF RADIO CIRCUITS

Design and Analysis of Complimentary Split Ring Resonator Backed Microstrip Transmission Line Using Equivalent Circuit Model[1]

P. Garg[a], * and P. Jain[a], **

[a] *Department of Electronics and Communication Engineering, Delhi Technological University, Shahbad Daulatpur, New Delhi, 110042 India*

**e-mail: priyankagarg_phd2k16@dtu.ac.in*

***e-mail: priyankajain@dtu.ac.in*

Received July 29, 2018

Abstract—This article presents a compact, low-profile Band-Stop Filter (BSF) designed using Complimentary Split Ring Resonator (CSRR). An equivalent circuit model is also presented along with the simplified mathematical approach to extract the parameters of the circuit model. This paper also presents the effect of variation in the dimensions of split rings on characteristics of BSF. The proposed BSF has a compact size of 27 × 20 mm^2 designed an FR-4 substrate with dielectric constant (ε_r) 4.3. The filter provides complete suppression of the band at 2.4 GHz. The design and circuit analysis of this metamaterial based filter is presented in terms of reflection coefficient, transmission coefficient and impedance curve.

Keywords: Bandstop filter (BSF), metamaterial, complementary split ring resonator (CSRR), microstrip line, resonator

DOI: 10.1134/S1064226918120069

INTRODUCTION

Bandstop filters form an essential element in various RF, microwave and other communication system for suppressing undesired frequency bands. Various band-stop filters have been proposed using planar technology [1, 2]. Although at some point, the design suffers from the drawback of large dimensions, use of metamaterials in such cases has proven to exhibit an exceptional reduction in dimensions.

Metamaterials are artificially engineered structures that provide extraordinary characteristics such as negative permittivity ε_r and permeability μ_r. Split ring resonators (SRR) have been shown to exhibit negative permeability around resonant frequency [3]. Complementary Split Ring Resonator (CSRR) is the counter image of SRR which shows negative permittivity for a particular band of frequencies [4]. CSRR etched on the ground plane at the back of a microstrip line is a well-known technique to realize filters. Such filters provide exceptional in-band and out-of-band performance. Liu et al. [5] presented a CSRR based high pass filter utilizing C-shaped coupling. An Open Complementary Split Ring Resonator (OCSRR) with 80% fractional bandwidth at central frequency 5 GHz is designed in [6]. A wideband bandpass filter with CSRR loaded microstrip transmission line is presented by Kim et al. [7] utilizing Evolution Strategy (ES) method for parameter optimization.

In this paper, a compact sized bandstop filter utilizing Complementary Split Ring Resonator (CSRR) is proposed. The proposed filter is designed to suppress the 2.4 GHz frequency band. Such filters can be extensively used in communication systems to avoid interference of original signal with other signals which may lead to error in output. The main objective of this paper is to demonstrate a methodology for selection of suitable equivalent circuit arrangement based on simulated results of the filter i.e. transmission coefficient and impedance curve, in order to study the electrical behavior of the design. Bonache et al. [8] presented the extraction of an equivalent circuit model, however using the methodology presented in this paper; we have found that the number of lumped elements used can be reduced further. A stepwise simplified mathematical approach for parameter extraction of the equivalent circuit is also presented. All the design and circuit simulations are performed using Computer Simulation Tomography Microwave Studio (CST-MWS) [9] and Computer Simulation Tomography Design Studio (CST-DS) [10]. Finally, the design is fabricated and measured.

[1] The article is published in the original.

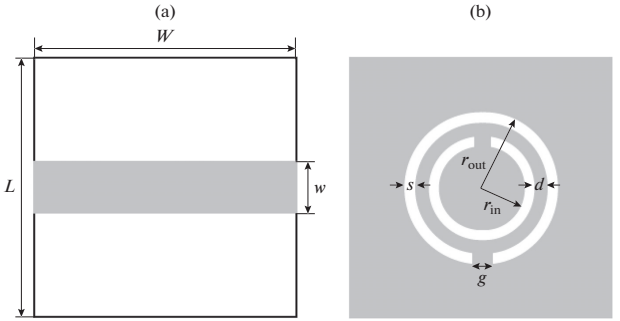


Fig. 1. Top microstrip line (a) and CSRR etched bottom ground plane (b).

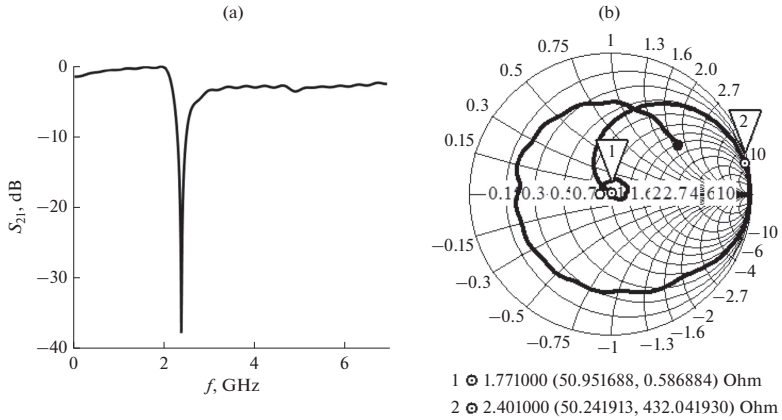


Fig. 2. The results of design simulation for frequency dependence of transmission coefficient (a) and impedance curve (b).

1. DESIGN AND CIRCUIT SIMULATION APPROACH

Figure 1 illustrates the geometry and dimensions of the proposed bandstop filter with a 50-Ω microstrip line on the top side and CSRR etched on the bottom ground plane. It is designed using an FR-4 dielectric substrate with relative permittivity (ε_r) 4.3, and thickness 1.6 mm. The proposed BSF has a compact size of only 27 × 20 mm^2. The dimensions of the microstrip line are chosen to deliver 50-Ω characteristic impedance.

The design parameters of proposed bandstop filter are given as: L = 20 mm, W = 27 mm, w = 3.1 mm, r_{out} = 5 mm, r_{in} = 2.85 mm, g = 0.6 mm, s = 0.7 mm and d = 0.75 mm.

In an SRR, each ring can be modeled as an inductor and the gap between the rings can be modeled as a capacitor. CSRR, being the dual of SRR, shows a complementary effect where the inductance is substituted by the capacitance of the disk and the gap capacitance is substituted by inductance between the slotted rings [11].

The circuit equivalent of the proposed BSF can be realized by using the transmission coefficient S_{21} and impedance characteristics, as shown in Figs. 2a, 2b respectively, obtained after design simulation. Since, the impedance curve shows inductive effect below resonant frequency and capacitive effect above it, so it is concluded that the circuit must contain a parallel res-

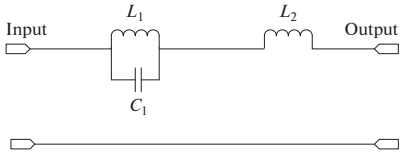


Fig. 3. Equivalent circuit of the proposed bandstop filter.

onant circuit. Further, S_{21} curve as shown in Fig. 2a, shows negative insertion loss (IL), given by,

$$\mathrm{IL} = 10\log_{10}|S_{21}(f)|^2. \quad (1)$$

Thus, the circuit equivalent of the simulated bandstop filter is a series connected parallel resonant circuit. The proposed bandstop filter shows maximum transmission at 1.771 GHz and minimum transmission at 2.401 GHz. Figure 3 shows the equivalent circuit obtained on the basis of simulated results.

Consider f_1 be the maximum transmission frequency, f_2 be the minimum transmission frequency and f_3 be the frequency at which IL is 3 dB such that $f_1 < f_3 < f_2$. The value of f_3 is obtained at the intersection of reflection and transmission coefficient S_{11} and S_{21} respectively, since,

$$|S_{11}|^2 + |S_{21}|^2 = 1 \quad (2)$$

and at the intersection point, $S_{11} = S_{21}$. Thus, $S_{11} = 1/\sqrt{2}$, a point of 3 dB IL.

The stepwise parameter extraction method, considering the circuit to be lossless, is summarized below.

(i) At the resonance condition of the parallel LC circuit i.e. f_2, S_{21} leads to zero. This gives the equation for C_1 in terms of L_1 and frequency f_2.

(ii) At the resonance condition of the whole tank circuit i.e. f_1, S_{11} leads to zero. This gives the relation for L_1 in terms of L_2 and frequencies f_1, f_2.

(iii) Finally, at 3 dB IL with frequency f_3, the third relation for L_2 in terms of frequency f_1, f_2 and f_3 is obtained.

The S-parameter matrix of the series connected parallel LC resonator is given by,

$$[S] = \begin{bmatrix} \frac{Y_0}{Y_0+2Y} & \frac{2Y}{Y_0+2Y} \\ \frac{2Y}{Y_0+2Y} & \frac{Y_0}{Y_0+2Y} \end{bmatrix},$$

where $Y_0 = 1/Z_0$, Z_0 is the characteristics impedance of the line which is 50-Ω and Y is the input admittance of the LC circuit.

The mathematical approach for parameter extraction using above steps is summarized below.

(i) Taking $S_{21} = 0$. Considering only parallel LC circuit (L_1,C_1), the input admittance is expressed as,

$$Y = \frac{1-\omega^2 L_1 C_1}{j\omega L_1}. \quad (3)$$

From the S-parameter matrix,

$$S_{21} = \frac{2Y}{Y_0+2Y} = \frac{2\left[\frac{1-\omega_2^2 L_1 C_1}{j\omega_2 L_1}\right]}{Y_0+2\left[\frac{1-\omega_2^2 L_1 C_1}{j\omega_2 L_1}\right]} = 0, \quad (4)$$

$$1-\omega_2^2 L_1 C_1 = 0. \quad (5)$$

On solving the above equation, we obtain,

$$C_1 = \frac{1}{4\pi^2 f_2^2 L_1}. \quad (6)$$

(ii) Equating S_{11} to zero,

$$S_{11} = \frac{Y_0}{Y_0+2Y} = \frac{Z}{Z+2Z_0} = 0. \quad (7)$$

Considering the input impedance of the overall tank circuit,

$$Z = \frac{j\omega_1 L_1}{1-\omega_1^2 L_1 C_1} + j\omega_1 L_2. \quad (8)$$

Substituting the value of Z from Eq. (8) to Eq.(7), we obtain,

$$\frac{j\omega_1 L_1}{1-\omega_1^2 L_1 C_1} + j\omega_1 L_2 = 0. \quad (9)$$

On solving, one can get,

$$L_1 + L_2 - \omega_1^2 L_1 L_2 C_1 = 0. \quad (10)$$

Substituting here the value of C_1 from Eq. (6), obtain

$$L_1 + L_2 - \frac{\omega_1^2}{\omega_2^2} L_2 = 0. \quad (11)$$

Solving for L_1, we get

$$L_1 = L_2\left[\frac{f_1^2}{f_2^2} - 1\right] = 0. \quad (12)$$

(iii) Now, the condition for 3 dB IL is given as,

$$|S_{21}| = \frac{2Y}{Y_0+2Y} = \frac{2Z_0}{Z+2Z_0} = \frac{1}{\sqrt{2}}, \quad (13)$$

$$\frac{2Z_0}{\sqrt{\left[\frac{j\omega_3 L_1}{1-\omega_3^2 L_1 C_1} + j\omega_3 L_2\right]^2 + 4Z_0^2}} = \frac{1}{\sqrt{2}}. \quad (14)$$

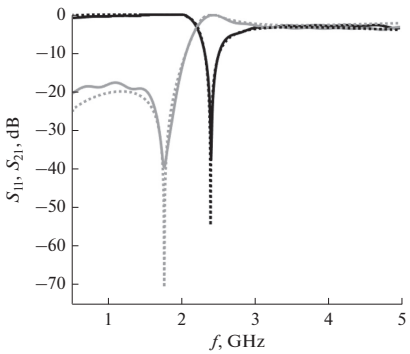


Fig. 4. Frequency dependences of reflection coefficients S_{11} (grey curves) and transmission coefficient S_{21} (blackcurves) obtained by design (solid curves) and circuit (dotted curves) simulation.

Now, on substituting the value of C_1 from Eq. (6) in Eq. (14) and solving, we get,

$$2Z_0\left(1-\frac{f_3^2}{f_2^2}\right) = \omega_3\left[L_1 + L_2\left(1-\frac{f_3^2}{f_2^2}\right)\right]. \quad (15)$$

Substituting the value of L_1 from Eq. (12) in Eq. (15) and solving, we obtain,

$$L_2 = \frac{Z_0\left(f_2^2 - f_3^2\right)}{\pi f_3\left(f_1^2 - f_3^2\right)}. \quad (16)$$

2. RESULTS AND DISCUSSIONS

From the results obtained after simulating the BSF design, we get f_1 = 1.771 GHz, f_2 = 2.401 GHz, f_3 = 2.217 GHz. Substituting all these values into Eqs. (6), (12), (16), we obtain the values of circuit parameters L_1, C_1, L_2 as 1.563 nH, 2.81 pF and −3.429 nH respectively. Using the values of L_1, C_1, L_2, the equivalent circuit of BSF is designed and simulated in CST design studio and results are obtained. Figure 4 shows the simulated reflection and transmission coefficient of the CSRR based BPF comparative to that of its circuit equivalent. Also, Figs. 5a and 5b simultaneously show the impedance curve of simulated design and circuit. Little mismatch in the impedance may be due to the presence of losses in the design. Both the results show that the circuit is a close approximation of the proposed BSF.

Now, it is important to note that the value of inductance L_2 is a negative quantity. It can be explained as the presence of an equivalent capacitance, say C_2' [12]

$$X_L = -j\omega L_2, \quad (17)$$

where negative sign indicates the negative nature of inductance L_2. Further rearranging the Eq. (17), we get,

$$X_L = \frac{1}{j\omega\left(\frac{1}{\omega^2 L_2}\right)}. \quad (18)$$

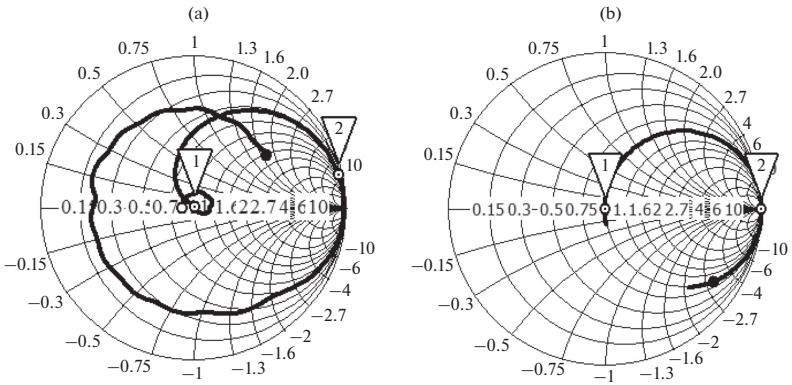


Fig. 5. Impedance curves obtained by design (a) and circuit (b) simulations.

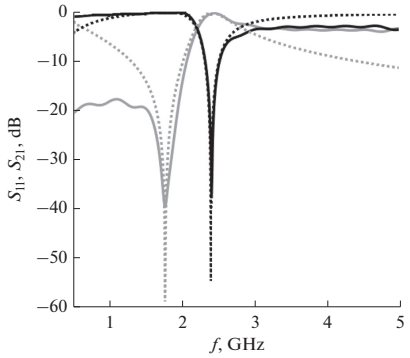


Fig. 6. Frequency dependences of reflection coefficients S_{11} (grey curves) and transmission coefficient S_{21} (black curves) obtained by design (solid curves) and circuit (dotted curves) simulation after replacing inductor L_2 shown in Fig. 3 by capacitor C_2'.

Thus,

$$C_2' = \frac{1}{\omega^2 L_2}. \quad (19)$$

Since there is two resonance phenomenon observed in the design, one is the parallel resonance occurring at f_2 and the other is the weak series resonance occurring at lower frequency f_1. At f_2, resonance is due to the parallel resonant circuit L_1 and C_1. At lower frequency, the parallel resonant circuit behaves as an inductor, say L_1'. On connecting C_2' in series with L_1' we get a weak series resonance at f_1.

From Eq. (19), C_2' at frequency f_1 is obtained as 2.35 pF. The simulated results after replacing the negative inductance L_2 by the equivalent positive capacitance C_2' is shown in Fig. 6.

3. PARAMETRIC ANALYSIS

Three dimensions are selected for parametric analysis that exhibit major variation in the characteristics of BSF. First one is the radius of the split rings. Figure 7a shows the variation in r_{out} keeping the dimensions s, d and g unchanged. It was observed that as r_{out} is increased from 4 mm to 6 mm, the resonant frequency shifts towards the lower side.

The second dimension for parametric analysis is the distance between the two rings d. Figure 7b shows variation in d keeping other parameters unchanged. As the parameter d is increased from 0.25 to 1.25 mm, the frequency again shifts to lower values.

Finally, variation in the split gap g is performed. Figure 7c shows as the split gap of the rings is increased from 0.2 to 1.8 mm, the frequency shifts towards higher values. Thus, suitable dimensions of the CSRR can be selected to obtain the desired resonant frequency as per the required specification.

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

Figure 8 shows the fabricated bandstop filter showing a microstrip transmission line on top and CSRR etched on the bottom ground plane. Keysight N9914A vector network analyzer is used to carry out the measurements of various parameters of the fabricated CSRR based BSF such as reflection coefficient, transmission coefficient, and 3-dB bandwidth. Figure 9 shows the S-parameters as observed on the Keysight N9914A vector network analyzer. Figure 10 simultaneously shows the measured; design simulated and circuit simulated values of S-parameters of the proposed BSF. The measured results show a 3-dB stopband from 2.113 to 3.088 GHz giving a 3-dB bandwidth of 0.975 GHz. The occurrence of ripples in the measured results may be due cable losses. The measured results match well with the simulated results.

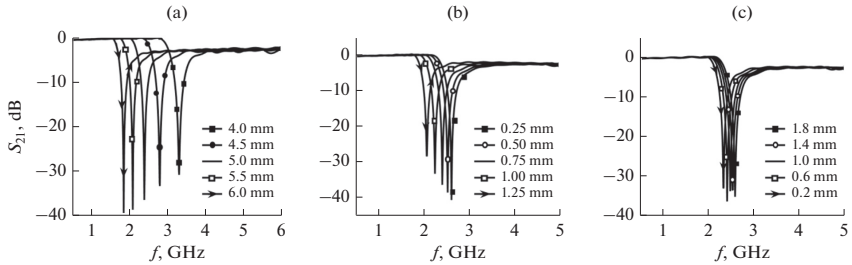


Fig. 7. Parametric analysis of frequency dependence of transmission coefficient S_{21}: r_{out} (a), d (b) and g (c).

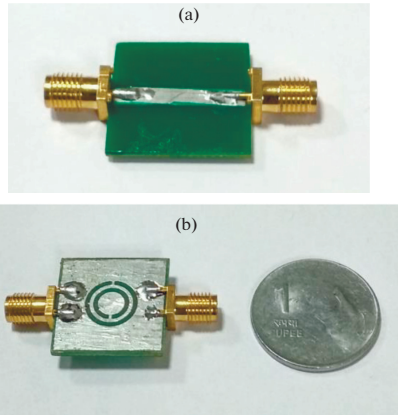


Fig. 8. Fabricated BSF: (a) top view, (b) bottom view. For comparison: the diameter of 1 Indian Rupee coin is approximately 22 mm.

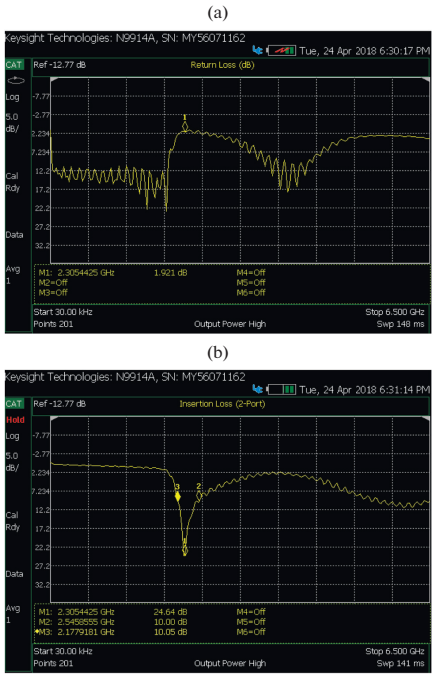


Fig. 9. Snapshots of results on Keysight N9914A vector network analyzer: (a) return loss, (b) insertion loss.

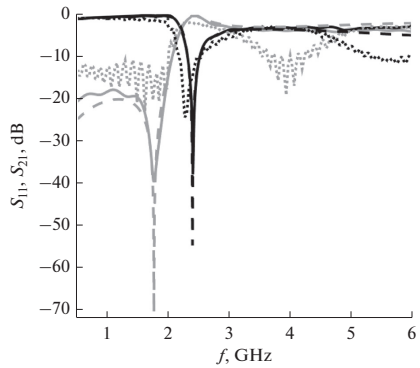


Fig. 10. Frequency dependences of reflection and transmission coefficients of proposed BSF: S_{11} (grey curves) and S_{21} (black curves). Solid curves are obtained by design simulation, dashed curves—by circuit simulation, dotted curves—measured values.

CONCLUSIONS

A complementary split ring resonator based bandstop filter has been designed and simulated. A stop band at the vicinity of 2.4 GHz is obtained. Based on the simulated results, the circuit arrangement is defined to be a series connected parallel resonant circuit followed by a series connected inductor. The simplified mathematical approach is described in detail to obtain the lumped circuit parameters. Only three lumped components are used to define the circuit model, thus reducing the circuit complexity. Furthermore, the parametric variation of the split rings is analyzed which displays frequency tuning capability by varying the dimensions of split rings. Finally, the filter is fabricated and measured to show fair agreement with the simulated results. The proposed bandstop filter shows good lower passband performance. The high-frequency passband shows large reflections which may be improved further.

REFERENCES

1. R. Habibi, C. Ghobadi, J. Nourinia, et al., Electron. Lett. **48**, 1483 (2012).
2. D. La, Y. Lu, S. Sun, et al., Microwave and Opt. Technol. Lett. **53**, 433 (2010).
3. J. B. Pendry, A. J. Holden, D. J. Robbins, et al., IEEE Trans. Microwave Theory Tech. **47**, 2075 (1999).

4. F. Falcone, T. Lopetegi, J. D. Baena, et al., IEEE Microwave and Wireless Compon. Lett. **14**, 280 (2004).
5. J. C. Liu, D. S. Shu, B. H. Zeng, et al., IET Microwaves, Antennas and Propag. **2**, 622 (2008).
6. A. Velez, F. Aznar, J. Bonache, et al., IEEE Microwave and Wireless Compon.Lett. **19**, 197 (2009).
7. K. T. Kim, J. H. Ko, K. Choi, et al., IEEE Trans. Magnetics **48**, 811 (2012).
8. J. Bonache, M. Gil, I. Gil, et al., IEEE Microwave and Wireless Compon. Lett. **16**, 543 (2006).
9. Computer Simulation Technology Microwave Studio (CST MWS). Available at http://www.cst.com.
10. Computer Simulation Technology Design Studio (CST DS). Available at http://www.cst.com.
11. J. D. Baena, J. Bonache, F. Martn, et al., IEEE Trans. Microwave Theory Tech. **53**, 1151 (2005).
12. A. Abdel-Rahman, A. K. Verma, A. Boutejdar, et al., IEEE Microwave and Wireless Compon. Lett. **14**, 136 (2004).

2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)

Detection of Ovarian Cancer using Organic Light Emitting Diodes

Shubham Negi[1], Poornima Mittal[2], Brijesh Kumar[3] and Anamika Bhatia[1]
[1]*Department of Electronics and Communication Engineering, Graphic Era (Deemed to be University) Dehradun*
[2]*Department of Electronics and Communication Engineering, Delhi Technological University, New Delhi*
[3]*Department of Electronics and Communication Engineering, Madan Mohan Malviya University of Technology, Gorakhpur, UP*
Shubham.negi.1992@ieee.org, brijesh_kumar@ieee.org; poornima2822@ieee.org

Abstract— **The modern world has been affected by a lot of diseases, and day by day new ones keeps on emerging. But no disease has threatened the mankind as much as cancer has. The reason behind it is that it has taken more human lives than any other disease. According to data of American Cancer Society, in USA alone half a million people will die because of cancer in 2018. Further this disease has a lot of different types which affects a whole lot of people varying in age and gender. Cancer even after all this is still a curable disease, but the problem with it that makes it so dangerous is late detection. If detected at early stages it can very easily be treated. Ovarian cancer is one of the categorization of cancer which has a very low survival rate often because of late detection and diagnosis. In this article a non invasive sensor mechanism has been proposed for detection of ovarian cancer. This method will result in a portable device that can be used for giving an early diagnosis of ovarian cancer and can easily be used at home.**

Keywords— Ovarian Cancer, Organic Light Emitting Diode (OLED), Organic Semiconductor (OSC).

I. Introduction

Cancer is a disease involving abnormal cell growth with the potential to invade and spread to other body parts. There are over 100 types of cancer that can affect humans. According to American Cancer Society's, Cancer "Facts and Figures 2018" report an estimated 1,735,350 new cancer cases will be diagnosed and 609,640 deaths will occur in United states alone [1]. Along with this the thing to focus is that Ovarian Cancer has got a special section in this year report because of its potential threat to women.

Even though ovarian cancer accounts for only a 2.5% of all cancer cases in women, but still it accounts for 5% death in women [1]. The reason for this being that 4 out of 5 ovarian cancer patients are diagnosed with advanced diseases. Its detection is tough and hence it is a research priority. Again according to American Cancer Society's data in 2018, approximately 22,240 new cases of ovarian cancer will be diagnosed and 14,070 deaths will occur in United States [1]. Early detection is the best way to beat it because in latter stages it often becomes incurable.

Organic Light Emitting Diodes (OLED) are the LED which uses organic semiconductor (OSC) materials for emitting light [2, 3]. These materials were earlier thought of as insulators [5-8], but with the advancement of technology and advent of thin film technology OSC showed conducting properties [4, 9-11]. Making use of this technology thin films of OSC are used for fabrication of OLED. OLEDs have a number of advantages such as huge color gamut, wide viewing angle, robust nature, flexibility etc [2, 3] and these advantages are due to presence of organic semiconductor materials [12-14]. Owing to these advantages a lot of new application of OLED are emerging like displays, sensors etc.

Sensor is a device or a transducer which detects a change in input and gives a corresponding output [2]. The input may be in the form of light, force, current voltage etc. Any device can be used as a detector if for a particular change in its input it gives a change in its response corresponding to the input.

In an article published in 2013, by M. Zvarik *et al.*, [15] showed that there was a difference in fluorescence in the urine samples of an ovarian cancer patients and that of a healthy women. The most significant difference was found in undiluted urine at a fluorescence emission wavelength from 400 nm to 460 nm, when excited at 310 to 390 nm. This can be used for easy and early detection of ovarian cancer with much ease and a low cost.

This paper has been consist of following five sections including the present introductory section. Thereafter, it is followed by section II where the principle for detection of ovarian cancer with the help of light has been presented. In section III, a brief has been presented as to how OLED can work as a sensor of light followed by proposed mechanism for detection of cancer with the help of OLED in section IV. Finally, in section V concluding remarks have been given.

II. Principle for Detection of Cancer by Light

Human urine contains a huge variety of biological compounds in the solution form such as metabolites, organic compounds and inorganic salts. The composition of these different solutes depends of different factors such as body metabolism, age hydration, dietary intake, and diseases. Thus urine analysis is widely used for detection of many diseases but it is generally a costly process. But recent studies have shown that urine consists of different intrinsic fluorophores, and modern fluorescence techniques can be used for testing and diagnosis of diseases in a patient.

A recent study by M. Zvarik *et al.* [15] showed that variation in urine autofluorescence can be used to differentiate between the healthy subjects and the ones suffering from ovarian cancer. In their study, they rigorously tested urine samples from a batch of healthy subjects and ovarian cancer patients.

978-1-5386-5002-8/18/$31.00 ©2018 IEEE

Their study revealed that there was a fivefold reduction in intensity of the peak when excited at 330 nm and has a peak emission at 420 nm for the cancer patients [16].

The study revealed that ovarian cancer patients possess weak fluorescence in the excitation region of 330-390 nm range as compared to samples from healthy person. Thus there was a red shift in the wavelength of the band for normal person at excitation of 330 nm to that for ovarian cancer patient at excitation of 370 nm [15]. This observed change in the band was most likely interpreted as significant reduction or even perishing of pyridoxic acid in the urine sample. The pyridoxic acid is the strongest fluorophore at excitation of 330nm and emission of 420 nm.

Based on the contrasts in the fluorescence response of the urine sample of healthy person and an ovarian cancer patient it becomes easy and cost effective to test a patient for ovarian cancer. To add on the test can be easily accomplished and to add on it is a non invasive technique.

Hitherto, method to detect the ovarian cancer with the help of fluorescence has been shown. The method used is no doubt a very good one as it is a non invasive method and shows good results. But the problem is that technique is yet dependent on a laboratory where a person has to go and the test has to be performed. It is well known that most of the people do not perform regular laboratory tests until and unless they face some major problem. So an easier solution for this problem would be to have a portable device which can be reused for testing of ovarian cancer at home. This device will be able to help give a preliminary diagnosis of ovarian cancer so that it can be detected at the early stages and can be treated more effectively.

III. OLED AS LIGHT DETECTOR/ SENSOR

OLED in the recent past have been used effectively as display element by major companies in their television sets and mobile phones. The reasons for this were the different advantages that are cited in the introduction. At the present time there is an extensive research going on to make flexible displays and mobile phones using OLED.

Therefore, the research in the field of OLED has been boosted recently. And it is because of this boost in the research arena that many different application of OLED are forthcoming day by day. It is being extensively being used in gas and chemical sensor application as well. Here in this section another application of OLED is being discussed i.e. OLED as a detector of light.

It is a well known fact that and light emitting diode 'LED' can be used for both emission as well as detection of light. Any LED can detect light which is smaller in wavelength than the wavelength of light that it emits. Following the same principle OLED has been analyzed for detection of LED.

Device has been analyzed with the help of state of the art tool Atlas by Silvaco [16]. It follows Poole and Frenkel mobility model [5, 7, 11, 16] for analysis of organic devices. The model is given as

$$\mu(E) = \mu_0 \exp\left[-\frac{\Delta}{KT} + \left(\frac{\beta}{KT} - \gamma\right)\sqrt{E}\right] \quad (1)$$

Where $\mu(E)$ = Field mobility

E = Electric Field

μ_0 = Null filed mobility

Δ = Activation Energy

β = Hole Poole Frenkel Factor = $q\sqrt{\frac{q}{\pi\varepsilon\varepsilon_0}}$

γ = Fitting Parameter

Further, Langevins Recombination Model [2, 16] is adopted for analysis of characteristics of OLED. This basic equation this model follows is

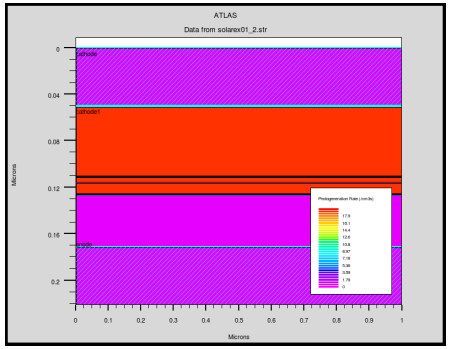
$$R_L(n,p) = r_l(x,y,t)\left(np - n_i^2\right) \quad (2)$$

Where r_l = Langevin recombination rate coefficient

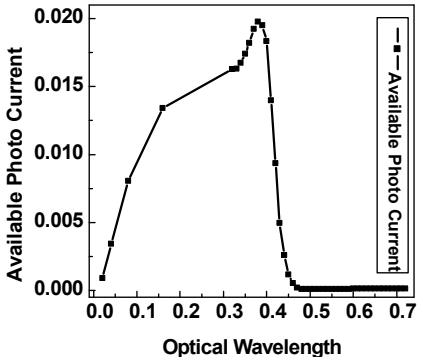
n_i = intrinsic concentration

n (p) = electron (hole) concentration

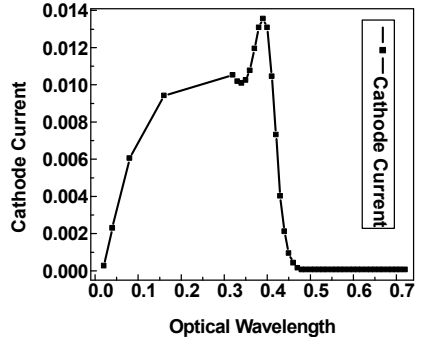
Based on this model a basic multilayered OLED was analyzed for detection of light. A light beam with varying light intensity was incident normally on the OLED. The intensity of light varied from 0 to 720 nm. Based on this the photon generation rate and photo current available was detected. The results of analysis are shown below.



(a)



(b)



(c)

Fig. 1 (a) Photon generation rate in the multilayered OLED, (b) Available photo current in the OLED, and (c) Cathode current generated by the OLED.

Thus the photo current and cathode current are generated only upto this level and thereafter there is a sharp dip in the both the current characteristics. Further, these two figures, Fig. 1 (b) and (c) shows that the current peaks around 430 nm and before this there is a gentle increase in current from 100 nm to 380 nm. It means that current varies according to the incident light and the variation in current to light is god enough to differentiate between different wavelength of light incident on it. Therefore, this phenomenon in OLED can be used for detection of cancer.

IV. PROPOSED MECHANISM FOR DETECTION OF LIGHT

The mechanism proposed for using OLED to detect the ovarian cancer is shown in Fig. 2 below.

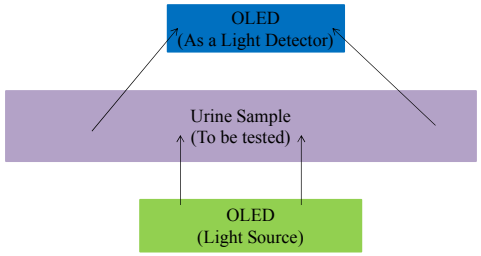


Fig. 2 Proposed mechanisms for detection of Ovarian Cancer

Fig. 2 shows that two OLEDs are required for detection cancer. One OLED (Green one at the bottom in Fig. 2) is used as a source of light for exciting the urine sample and thereafter second OLED (Blue one at the top of Fig. 2) is used for detection of fluorescence from the different components of urine sample. If the sample is from a healthy person, then the current from the OLED will be high as the emission will be in the range of 420-450 nm. Whereas if the person is suffering from ovarian cancer the generated current in detector OLED will be small as emission will be in the range of 350 nm. Because of this difference in emission spectra the detected current in our device will vary as can be seen in Fig. 1 (b) and (c). For emission in 420- 450 nm spectra current is in the range of 20 mA whereas for emission in 350 nm spectra the current will be in the range of 10 mA spectra. As a result of this diagnosis of ovarian cancer can be made possible.

V. CONCLUSION

In this paper a portable non invasive mechanism has been proposed for the detection of ovarian cancer with the help of OLED. Here the OLED is used for both: as a light source and as a sensor for detection of light. The fact that urine sample of healthy person and that of patient suffering from ovarian cancer has different fluorescence spectra has been exploited here for preliminary diagnosis of ovarian cancer. OLED that has been used as light detector is capable of detecting and differentiating between the two different emission spectra and thus it generates the current based on the two emission spectra. Healthy person's urine sample shows a higher fluorescence and thus the current generated is higher. On the other hand, patient suffering from ovarian cancer shows lower fluorescence and thus the current generated is smaller. Based on these two currents the device can be calibrated and a portable device can be made for detection of ovarian cancer easily.

REFERENCES

[1] American Cancer Society. Cancer Facts & Figures 2018. Atlanta: American Cancer Society; 2018.

[2] B. Kumar, B. K. Kaushik, and Y. S. Negi, "Organic Thin Film Transistors: Structure, Models, Materials, Fabrication, and Applications: A review," Polymer Review, vol. 54:1, pp. 33-111, Sept. 2013.

[3] S. Negi, P. Mittal, and B. Kumar, "Impact of different layers on performance of OLED," *Microsystem Technologies*, vol. *24, no.* 12, pp. 4981-4989, 2018.

[4] B. Kumar, B. K. Kaushik, and Y. S. Negi, "Design and analysis of noise margin, write ability and read stability of organic and hybrid 6-T SRAM," *Microelectronics Reliability*, vol. 54, no. 12, Dec. 2014, pp. 2801–2812, (Published by Elsevier, SCI, and Thomson Reuters Impact Factor: 1.371).

[5] B. Kumar, B. K. Kaushik, and Y. S. Negi, "Static and dynamic characteristics of dual gate organic TFT based NAND and NOR circuits," *Journal of Computational Electronics*, vol. 13, no. 3, pp. 627-638, Sept. 2014, (Published by Springer, SCI, Thomson Reuters Impact Factor: 1.526).

[6] B. Kumar, B. K. Kaushik, and Y. S. Negi, "Analysis of electrical parameters of organic thin film transistors based on thickness variation in semiconducting and dielectric layers," *IET Circuits, Devices & Systems*, vol. 8, no. 2, pp. 131–140, March, 2014, (Published by IET/Available on IEEE Explore, SCI, Thomson Reuters Impact Factor: 1.092).

[7] B. Kumar, B. K. Kaushik, Y. S. Negi, and V. Goswami, "Single and dual gate OTFTs based robust organic digital design," *Microelectronics Reliability*, vol. 54, no. 1, pp. 100–109, Jan. 2014, (Published by Elsevier, SCI, Thomson Reuters Impact Factor: 1.371).

[8] B. Kumar, B. K. Kaushik, and Y. S. Negi, "Perspectives and challenges for organic thin film transistors: Materials, devices, processes and applications", *Journal of Material Science: Materials in Electronics*, vol. 25, no. 1, pp. 1–30, Jan. 2014, (Published by Springer, SCI, Thomson Reuter, Impact Factor: 2.019).

[9] B. Kumar, B. K. Kaushik, and Y. S. Negi, "Static and dynamic analysis of organic and hybrid inverter circuits," *Journal of Computational Electronics*, vol. 12, no. 4, pp. 765–774, Dec. 2013, (Published by Springer, SCI, Thomson Reuters Impact Factor: 1.526).

[10] V. Goswami, B. Kumar, B. K. Kaushik, Y. S. Negi, and K. L. Yadav, "Analysis of static and dynamic performance of organic inverter circuits based on dual and single gate OTFTs," *IET Circuits, Devices & Systems*, vol. 7, no. 6, pp. 345–351, Nov. 2013, (Published by IET/Available on IEEE Explore, SCI, Thomson Reuters, Impact Factor: 1.092).

[11] B. Kumar, B. K. Kaushik, Y. S. Negi, S. Saxena, and G. D. Varma, "Analytical modeling and parameter extraction of top and bottom contact structures of organic thin film transistors," *Microelectronics Journal*, vol. 44, no. 9, pp. 736–143, Sept. 2013, (Published by Elsevier, SCI, Thomson Reuters, Impact Factor: 1.163).

[12] B. Kumar, B. K. Kaushik, and Y. S. Negi, "Modeling of top and bottom contact structure organic field effect transistors," IEEE, *Journal of Vacuum Science and Technology B: Microelectronics and Nanometer Structures*, vol. 31, no. 1, pp. 012401-1–012401-7, Jan. 2013, (Published by AVS/ Available on IEEE Explore, SCI, Thomson Reuters, Impact Factor: 1.358).

[13] B. Kumar, B. K. Kaushik, and Y. S. Negi, "Analysis of electrode thickness variation effect on performance parameters of polymer thin film transistor with device simulation," *Int. J. Advanced Intelligence Paradigms,* vol. 5, no. 1/2, pp. 3–15, Jan. 2013, (Published by Inder Science Publisher, Indexed by Scopus, Elsevier).

[14] P. Mittal, B. Kumar, B. K. Kaushik, Y. S. Negi and R. K Singh, "Channel length variation effect on performance parameters of organic field effect transistors, *Microelectron. Journal,* vol. 43, no. 12, pp. 985–994, Dec. 2012, (Published by Elsevier, SCI, Thomson Reuters and Impact Factor: 0.924).

[15] M. Zvarik, D. Martinicky, L. Hunakova, I. Lajdova, I., and L. Sikurova, "Fluorescence characteristics of human urine from normal individuals and ovarian cancer patients," Neoplasma, vol. 60, no.5, pp. 533-537, 2013.

[16] ATLAS User's Manual Device Simulation Software. Silvaco International, Santa Clara, 2010.

Accepted Manuscript



Electrochemical paper based cancer biosensor using iron oxide nanoparticles decorated PEDOT:PSS

Saurabh Kumar, Mohammad Umar, Anas Saifi, Suveen Kumar, Shine Augustine, Saurabh Srivastava, Bansi D. Malhotra

PII: S0003-2670(19)30017-0

DOI: https://doi.org/10.1016/j.aca.2018.12.053

Reference: ACA 236494

To appear in: *Analytica Chimica Acta*

Received Date: 28 October 2018

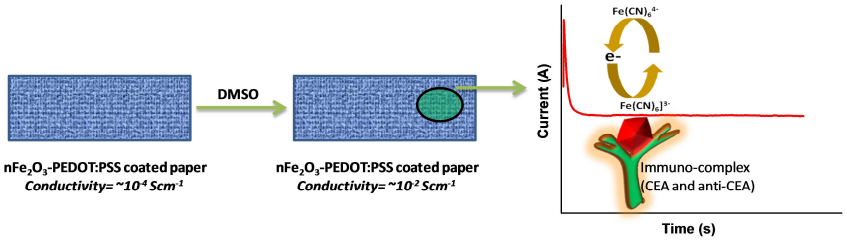
Revised Date: 16 December 2018

Accepted Date: 20 December 2018

Please cite this article as: S. Kumar, M. Umar, A. Saifi, S. Kumar, S. Augustine, S. Srivastava, B.D Malhotra, Electrochemical paper based cancer biosensor using iron oxide nanoparticles decorated PEDOT:PSS, *Analytica Chimica Acta*, https://doi.org/10.1016/j.aca.2018.12.053.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT

ELECTROCHEMICAL PAPER BASED CANCER BIOSENSOR USING IRON OXIDE NANOPARTICLES DECORATED PEDOT:PSS

Saurabh Kumar[1,2†], Mohammad Umar[1†], Anas Saifi[1], Suveen Kumar[1,3], Shine Augustine[1], Saurabh Srivastava[1,4*], Bansi D Malhotra[1*]

[1]Nanobioelectronics Laboratory, Department of Biotechnology, Delhi Technological University, Shahbad Daulatpur, Delhi 110042, India

[2]Centre for Nano Science and Engineering (CeNSE), Indian Institute of Science, Bangalore-560012, India

[3]Department of Chemistry, University of Delhi, Delhi 110007, India

[4]Department of Applied Science & Humanities, Rajkiya Engineering College, Ambedkar Nagar, UP 224122, India

ACCEPTED MANUSCRIPT

ABSTRACT: We report results of the studies relating to the fabrication of a label-free, flexible, light weight and disposable conducting paper based immunosensing platform comprising of poly(3,4-ethylenedioxythiophene):poly(styrenesulfonate) and nanostructured iron oxide (nFe_2O_3@PEDOT:PSS) nanocomposite for detection of carcinoembryonic antigen (CEA), a cancer biomarker. The effect of various solvents such as sorbitol, ethanol, propanol, n-methyl-2-pyrrolidone (NMP) and dimethyl sulfoxide (DMSO) on the electrical conductivity of Whatman filter paper (WP) modified with nFe_2O_3@PEDOT:PSS/WP was investigated. The electrical conductivity of the PEDOT:PSS/WP electrode was found to be enhanced by two orders of magnitude (from 6.8 x 10^{-4} to 1.92 x 10^{-2} Scm^{-1}) after its treatment with DMSO. Further, nFe_2O_3 doped PEDOT:PSS/WP electrode exhibited the electrical conductivity as 2.4 x 10^{-2} Scm^{-1}. Besides this, the incorporation of iron oxide nanoparticles (nFe_2O_3) into PEDOT:PSS/WP resulted in improved electrochemical performance and signal stability. This nFe_2O_3@PEDOT:PSS/WP based platform was used for immobilization of the anti-carcinoembronic antigen (anti-CEA) protein for quantitative estimation of cancer biomarker (CEA). The results of electrochemical response studies revealed that this conducting paper based immunoelectrode had a sensitivity of 10.2 $\mu Ang^{-1}mLcm^{-2}$ in the physiological range (4-25 $ngmL^{-1}$) and shelf life of 34 days. Further, the proposed immunoelectrode was validated with conventional ELISA for the detection of CEA in serum samples of cancer patients.

Keywords: PEDOT: PSS, Iron oxide nanoparticles, Electrochemical paper, Biosensor.

Corresponding Authors
*Email: bansi.malhotra@gmail.com (B.D. Malhotra), saurabhnpl@gmail.com (Saurabh Srivastava)

Author contributions
Ŧ These authors contributed equally

ACCEPTED MANUSCRIPT

1. INTRODUCTION

Paper has been used for a large number of quantitative analysis since ancient times. However, the qualitative or semi-quantitative results obtained using traditional paper-based detection methods have been reported to have limited applications [1-2]. This has been attributed to the fact that these detection methods are often dependant on the visual observations. With advancement in techniques and exploration of different materials, paper has recently been re-discovered as a sensor substrate which provides a platform for the development of paper based analytical devices [3-4]. In this context, fabrication of the conducting paper based biosensors has recently aroused much interest. Numerous materials such as metals [5], metal oxides [6], carbon nanomaterials [7-8], and conducting polymers [9-10] can be employed for fabrication of the conducting paper. Incorporation of inorganic materials such as metal and metal oxide into paper may result in improved electrical performance. However, complex processing, limited flexibility and high material cost have led to limited applications towards the fabrication of flexible conducting paper. The preparation of carbon nanomaterials necessitates tedious fabrication steps and these are known to be insoluble in most solvents making it difficult to prepare uniform films over large areas. In this context, conducting polymers are getting increased attention towards the fabrication of conducting paper as these offer good conductivity, solution processability, mechanical flexibility and are low cost [11]. Conducting polymers have conjugated π-electrons in their backbone structure and overlapping of their partially filled molecular orbitals permits the formation of delocalized molecular wave function. This allows electrons to move freely throughout the lattice [12] leading to improved electronic and optical properties [13]. Poly(3,4ethylenedioxythiophene):poly(styrenesulfonate) (PEDOT:PSS) conducting polymer has been found to be a suitable candidate for fabrication of conducting paper owing to its uniform

coating on paper while retaining the flexibility and stability [11]. Besides this, it manifests relatively high conductivity, improved electrochemical and thermal stabilities, greater transparency and good quality films can be easily coated onto desired substrates by conventional solution processing methods [14-18].

The nanostructured iron oxide has been predicted to be a promising immobilization matrix due to its fast electron-transfer kinetics, strong adsorption capacity, high surface-to-volume ratio, biocompatibility and non-toxic nature. This nano-structured material may offer suitable microenvironment for biomolecules immobilization leading to improved biosensing characteristics [19]. Moreover, a composite of PEDOT:PSS and nFe_2O_3 may provide synergistic effect resulting in the development of a flexible, lightweight, low cost and disposable biosensor. This organic-metal oxide nanohybrid (nFe_2O_3@PEDOT:PSS) modified conducting paper may facilitate improved conduction of electrons and ionic charge carriers within the hierarchical structure of cellulose paper.

Cancer is currently a major cause of mortality and medical threat worldwide. About 14 million new cases and 8.2 million cancer related deaths were reported in 2012 [20]. Carcinoembryonic antigen (CEA) is an important biomarker that can be correlated with the on-set of colorectal, ovarian, lung, pancreatic, breast and liver cancer [21]. According to a WHO report, these cancer are responsible for more than half of all deaths occurring annually [22]. Therefore, quantification of the CEA level before and after cancer treatment may facilitate early detection of this dreadful disease. Many conducting polymer based electrochemical sensors have been reported in literature for CEA detection. Zhang *et al.* deposited polyaniline/polypyrrole-silver dendrites composite on ITO electrode for CEA detection using electrochemiluminescence (ECL) technique and achieved a linear detection range from 0.001-100 $ngmL^{-1}$ in logarithmic

scale [23]. Similarly an ECL based sensor was fabricated using poly(5-formylindole)/reduced graphene oxide nanocomposite (P5FIn/erGO) and obtained linear range of 0.0001-10 $ngmL^{-1}$ in logarithmic scale [24]. In another work Nie *et al.* deposited poly(5-formylindole)/electrochemically reduced graphene oxide (P5FIn/erGO) and gold nanoparticle composite onto ITO electrode [25]. The electrochemical response was found to be linear in the range, 0.0005 to 50 $ngmL^{-1}$. Zhao *et al.* reported glassy carbon electrode for CEA detection using nanocomposite of electrochemically reduced graphene oxide, gold nanoparticles (AuNPs) and poly(indole-6-carboxylic acid) [26]. These methods require additional labelling steps and/or are based on expensive conventional electrodes (ITO or glassy carbon). Due to low cost, flexible and disposable nature, conducting paper based electrochemical biosensors have been found to have many applications in smart point-of-care (POC) devices. Kumar *et al.* fabricated reduced graphene oxide modified conducting paper for amperometric detection of CEA [11]. This conducting paper platform had narrow linear detection range (2–8 $ngmL^{-1}$) and short shelf life (21 days). We report a facile method for the fabrication of electrochemical paper biosensor for CEA biomarker detection using iron oxide nanoparticles decorated PEDOT:PSS. The proposed electrochemical paper biosensor exhibited efficient linear detection range, high sensitivity and long term stability.

2. EXPERIMENTAL SECTION

2.1. Materials.

PEDOT:PSS (1.3 wt%) and biological reagents [CEA, anti-CEA, bovine serum albumin (BSA)] were procured from Sigma Aldrich. DMSO and Whatman filter paper Grade 1 were purchased from High Purity Laboratory Chemicals, India and GE healthcare, UK, respectively. Buffer and

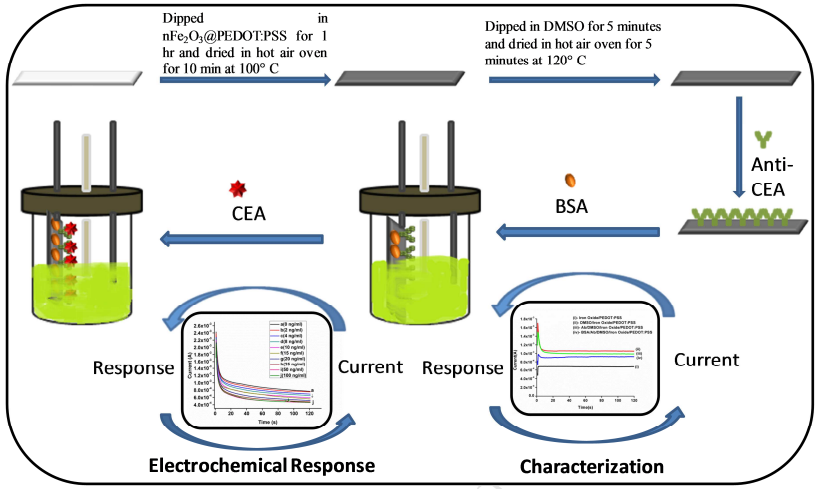
desired solutions were prepared using Milli-Q water, oleic acid capped iron oxide nanoparticles (nFe_2O_3) were received as a gift.

2.2. Fabrication of electrochemical paper electrode.

An optimized amount, 10 mg of nFe_2O_3 was dispersed into 10 mL of PEDOT:PSS solution via ultrasonication method. Further, Whatman filter paper (WP, 1 cm x 3 cm) #1 was dipped for 1 h into nFe_2O_3@PEDOT:PSS solution followed by drying in hot air oven at 100˚C for 10 minutes. Thereafter, the dried paper electrode (nFe_2O_3@PEDOT:PSS/WP) was dipped in DMSO solution for about 5 minutes to obtain enhanced conductivity of the paper electrode and was oven-dried at 120˚C for 5 minutes. The DMSO treated nFe_2O_3@PEDOT:PSS/WP electrode was termed as the electrochemical paper (EP) electrode.

2.3. Antibody Immobilization.

The monoclonal carcinoembryonic antibodies (anti-CEA) were immobilized onto EP electrode via physiological adsorption. In this process 20µl solution of anti-CEA (1000 µgmL^{-1}) in PBS (pH 7.4) was drop-cast over the surface (1x1 cm^2) of EP electrode and incubated for 6 h. It was then rinsed with PBS to remove any unbound antibodies after which the anti-CEA/EP electrode was treated with 0.1 % BSA to block non-specific active sites. In this process, 20µl solution of 0.1 % BSA was drop-cast over the electrode surface, was kept undisturbed for 2 h and then rinsed with PBS. Finally, the BSA/anti-CEA/EP electrode was stored in a refrigerator at 4 °C when not in use. During electrochemical measurements, the EP electrode was dipped into the electrochemical cell up to 1cm. Therefore the detecting area of EP electrode was 1x1cm^2. A schematic of the fabrication electrochemical paper sensor is shown in scheme 1.



Scheme 1: Schematic representation of fabrication of electrochemical paper based cancer biosensor.

2.4. Characterization.

The prepared nanocomposite (nFe_2O_3@PEDOT:PSS) was characterized by Tecnai G2 30, ultratwin transmission electron microscopy (TEM) and by Bruker D-8 Advance, X-ray diffraction spectroscopy (Cu Kα, l = 1.5406 Å). Conductivity measurements on the various modified electrochemical papers were done using four points probe technique containing a low current source, and digital micro-voltmeter (SES Instruments, India). Surface morphology of various modified electrochemical papers electrode was studied by an SEM (Hitachi S-3700N) and X-ray photoelectron spectroscopy (XPS, Scientia ESCA 200). The electrochemical studies were performed on an Autolab Potentiostat/Galvanostat (Metrohm, Netherland) using a three electrode system with electrochemical paper electrode as working electrode, platinum wire as the

counter electrode and silver-silver chloride (Ag/AgCl) as reference electrode in PBS (50 mM, pH 7.0) containing 5 mM $[Fe(CN)_6]^{3-/4-}$ redox species

3. RESULTS AND DISCUSSION

3.1 Transmission electron microscopy (TEM)

Figure 1(a) shows the transmission electron microscopy (TEM) image of nFe_2O_3 (oleic acid caped) that are nearly spherical in shape having a particle size of ~10–12 nm. However, a few of these are non-spherical or elliptical. In the TEM image of nFe_2O_3@PEDOT:PSS composite (figure 1(b)), the clusters of nFe_2O_3 having size of a few hundred nanometres were visible over the PEDOT: PSS substrate. However, these were not aggregated (figure 1(c)) and were found to be embedded and encapsulated in PEDOT:PSS polymer matrix. It appears that oleic acid capped nFe_2O_3 carrying negative charge and PEDOT being positive moieties resulted in the electrostatic adsorption of amorphous PEDOT polymer over nFe_2O_3 [27].

3.2 X-ray diffraction (XRD) studies

Figure 1(d) shows XRD patterns obtained for (i) PEDOT:PSS, (ii) DMSO treated PEDOT:PSS and (iii) iron oxide nanoparticles doped PEDOT:PSS powder sample. The X-Ray diffraction pattern of the PEDOT:PSS showed a broad peak at 2θ=24.70º [28]. This diffraction peak perhaps arose due to PEDOT thiophene ring (π-π) stacking [29]. The DMSO treated PEDOT:PSS showed the peak shift from 2θ=24.70º to 2θ=24.84º which corresponded to the increase in interchain planar ring stacking and a decrease in intensity was observed due to rearrangement of the crystalline domain orientation [30]. It may be noted that PEDOT is known to be highly conducting and is surrounded by non-conducting PSS molecule. After treatment with DMSO there is removal of some PSS chains from the PEDOT:PSS film [31]. That may perhaps cause conformational change in the PEDOT:PSS polymer structure. Therefore, the observed change in

XRD angle (24.70º to 24.80º) indicated small conformational change and was not related to any variation of the PEDOT:PSS polymer structure. XRD of nFe_2O_3@PEDOT:PSS showed a peak shift from 24.7 to 26.1 arising due to nFe_2O_3 entrapment in the PEDOT:PSS polymer chain. Additional peaks seen at 2θ =30.65, 35.70, 53.28, 57.43, 63.03 corresponding to the (2 2 0), (3 1 1), (4 2 2), (5 1 1), (4 4 0) planes were well indexed with the standard XRD pattern of nFe_2O_3 (JCPDS No. 39–1346) [28].

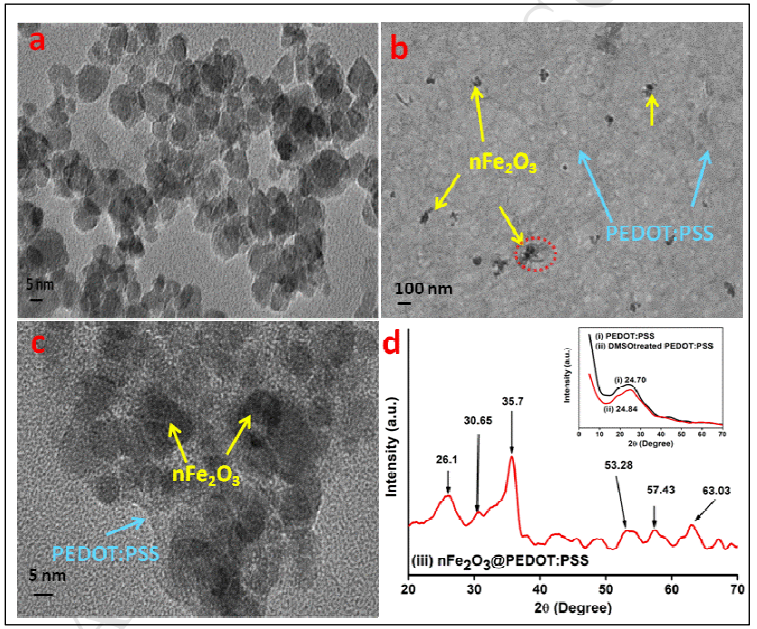


Figure 1: TEM image of (a) nFe_2O_3, (b) nFe_2O_3@PEDOT:PSS composite, (c) enlarge view of nFe_2O_3@PEDOT:PSS composite and (d) XRD pattern of nFe_2O_3@PEDOT:PSS and the inset showing XRD of (i) PEDOT:PSS and (ii) DMSO treated PEDOT:PSS.

3.3 Scanning electron microscopy (SEM)

The surface morphology of the electrochemical paper was examined using a scanning electron microscope (SEM). In Figure 2(a) cellulose fibers of Whatman filter paper (WP) were clearly seen and in Figure 2(b) PEDOT: PSS was found to be uniformly coated on WP which had perhaps smoothened the cellulose fiber. The PEDOT: PSS was largely adsorbed in/on the cellulose fibers network via capillary action. Figure 2(c) showed small spherical moieties of nFe_2O_3 embedded in the polymer structure of PEDOT: PSS/WP (electrochemical paper, EP). Further, shiny appearance on the electrochemical paper electrode was observed (Figure 2d) because of the static charge build-up arising due to non-conducting nature of the CEA antibodies which supported the immobilization of CEA antibodies. The immobilized antibodies filled up the pores and small interfiber air spaces of cellulosic network present in the nFe_2O_3/EP.

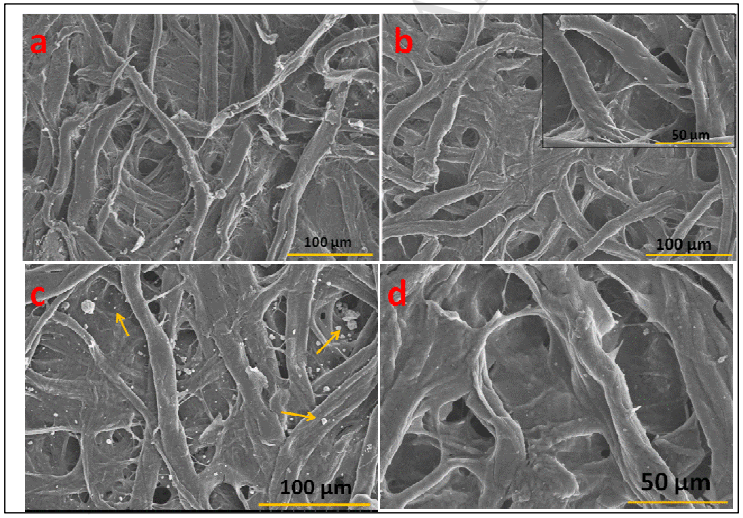


Figure 2: SEM image of (a) Whatman filter paper (WP) (b) PEDOT: PSS/WP, (c) nFe_2O_3/PEDOT: PSS/WP (EP), (d) anti-CEA/EP.

3.4 Electrical conductivity studies

The electrical conductivity of PEDOT:PSS coated paper (PEDOT:PSS/WP) was found to be ~6.8 x 10^{-4} Scm^{-1}. Further, we treated the PEDOT:PSS/WP electrode with different solvents and investigated its electrical conductivity (Table 1). When PEDOT:PSS/WP electrode was treated with DMSO , its electrical conductivity was improved greatly (6.8 x 10^{-4} to 1.92 x 10^{-2} Scm^{-1}) As a result, DMSO treated PEDOT:PSS/WP electrode was chosen for further studies. The reason for the increase in conductivity after DMSO treatment was attributed to the charge interaction between PEDOT and PSS which caused the reorientation of the polymer enhancing the mobility of charge carriers [30]. DMSO being an aprotic solvent formed bond with hydrophobic and conducting PEDOT conjugating polymer and PSS was washed away with distilled water [31]. The PEDOT:PSS doped with iron oxide nanoparticles after treatment with DMSO (electrochemical paper) showed electrical conductivity of 2.4 x 10^{-2} Scm^{-1} which was nearly similar to DMSO treated PEDOT:PSS/WP.

Table 1: Effect of various solvents on the electrical conductivity of conducting paper.

S. No.	Sample	Conductivity (S/cm)
1	PEDOT:PSS coated on paper, CP	6.84 x 10^{-4}
2	CP dipped in N-methyl-2-pyrrolidone	6.9 x 10^{-4}
3	CP dipped in Sorbitol	1.3 x 10^{-3}
4	CP dipped in Propanol	3.6 x 10^{-4}
5	CP dipped in Ethanol	9.7 x 10^{-4}
6	CP dipped in DMSO, DMSO@CP	1.92 x 10^{-2}
7	nFe_2O_3/DMSO@CP	2.40 x 10^{-2}

The XPS studies were carried out to investigate the effect of DMSO treatment on PEDOT:PSS/WP electrode. The S2p core level spectra of PEDOT:PSS and DMSO treated PEDOT:PSS showed two specific peaks(figure 3(a)) . The S2p peak present at lower binding

energy (163.4 and 165.8 eV) belonged to PEDOT and S2p peak at higher binding energies (167.4 and 168.8 eV) corresponded to PSS. The doublet present in each S2p peak was due to spin orbit splitting ($2p_{3/2}$ and S $2p_{1/2}$) [32-33]. Besides this, it was found (figure 3(a)), that the contribution to the peak intensity of S2p from the PEDOT moiety was significantly increased after the DMSO treatment. This was due to exclusion of PSS ions upon treatment with DMSO. Further, the peak intensity ratio of PEDOT/PSS electorde increased from 0.52 to 0.69 before and after treatment with DMSO solvent, respectively. This increase was attributed to the increased contribution of the PEDOT component to the signal. Moreover, a shift in the peak of PEDOT spectra from 165.80 eV to 165.20 eV indicated that rearrangement in PEDOT structure had occurred.

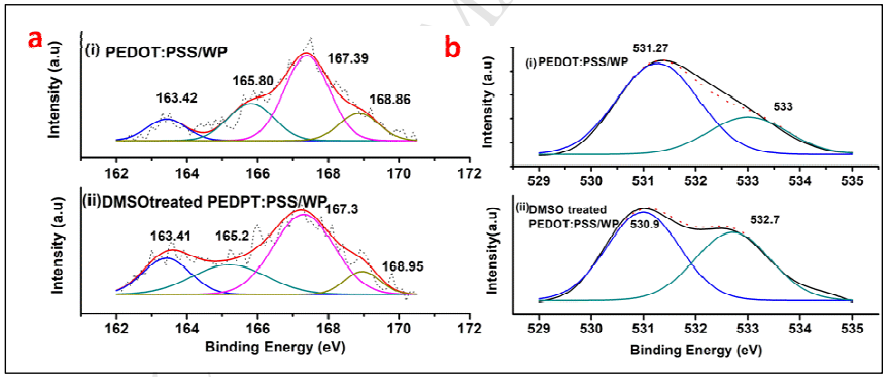


Figure 3: XPS spectra deconvulation of a) S2p and b) O1s of PEDOT:PSS/WP and DMSO treated PEDOT:PSS/WP.

Figure 3(b) is the O1s spectra of PEDOT:PSS/WP obtained before and after its treatment with DMSO .It revealed peaks at around 533.0 and 531.27 eV that were assigned to PEDOT and PSS moieties, respectively. The peak at 531.27 eV was due to oxygen atoms bound with the sulfur atoms via double bond in the PSS chain .The peak obtained at 533.0 eV corresponded to the oxygen atom of PEDOT chain [32]. The peak intensity due to the PEDOT moietis was found to be increased after treatment its with DMSO. PEDOT/PSS peak intensity ratio was found to increase from 0.39 to 0.79 after the DMSO treatment, indicating removal of PSS from the conducting polymer.

3.5 Flexibility study

The flexible nature of electrochemical paper electrode was investigated by folding of the electrode in -180° to +180° fashion (figure 4(a)) wherein positive and negative angles indicated folding of the electrode in upward and download direction, respectively. The result was plotted as relative conductivity [ratio of measured conductivity (σ_m) to the initial conductivity (σ_o), σ_m/σ_o] against the folding angle. Figure 4(b) shows variation in relative conductivity of electrochemical paper electrode up to 17% by folding. It can be concluded that the electrodes were flexible and there was less variation in conductivity after folding.

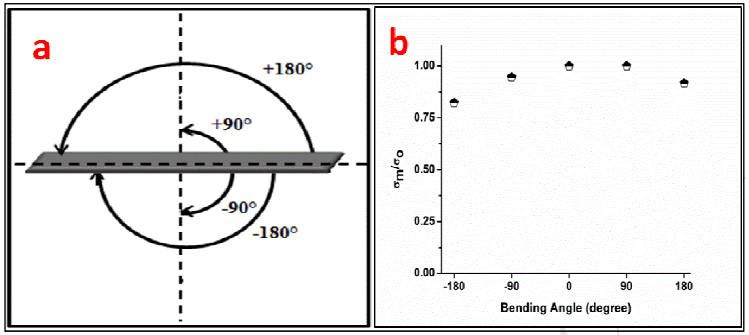


Figure 4: (a) Schematic diagram of electrode folding at different angles. **(b)** Shows plot of relative conductivity vs. bending angle.

3.6 Electrochemical studies

The electrochemical studies of (i) nFe_2O_3/PEDOT:PSS/WP, (ii) DMSO treated nFe_2O_3/PEDOT:PSS/WP, [electrochemical paper, EP] (iii) anti-CEA immobilized over electrochemical paper [anti-CEA/EP] and (iv) BSA/anti-CEA/EP electrodes were investigated via chronoamperometric technique at regular interval of 0.1 s at a potential of 2V (figure 5a). As shown in the figure, the electrochemical current of nFe_2O_3/PEDOT:PSS/WP electrode was observed to be ~0.65 mA. However, after treatment with DMSO the electrochemical current increased to 1.06 mA. This increase in electrochemical current was attributed to the ejection of insulating PSS ions from nFe_2O_3/PEDOT:PSS/WP electrode surface, resulting in enhanced permeability of $[Fe(CN)_6]^{3-/4-}$ redox couple [31]. After immobilization of anti-CEA onto the EP electrode the electrochemical (EC) current was found to be decreased to 0.98 mA. This decrease in EC could be due to hindrance in electron transport at the electrode surface caused by the CEA antibodies. On treatment of anti-CEA/EP with BSA a decrease in EC current (0.92 mA) was observed due to insulating nature of BSA which hindered the electron transport [11].

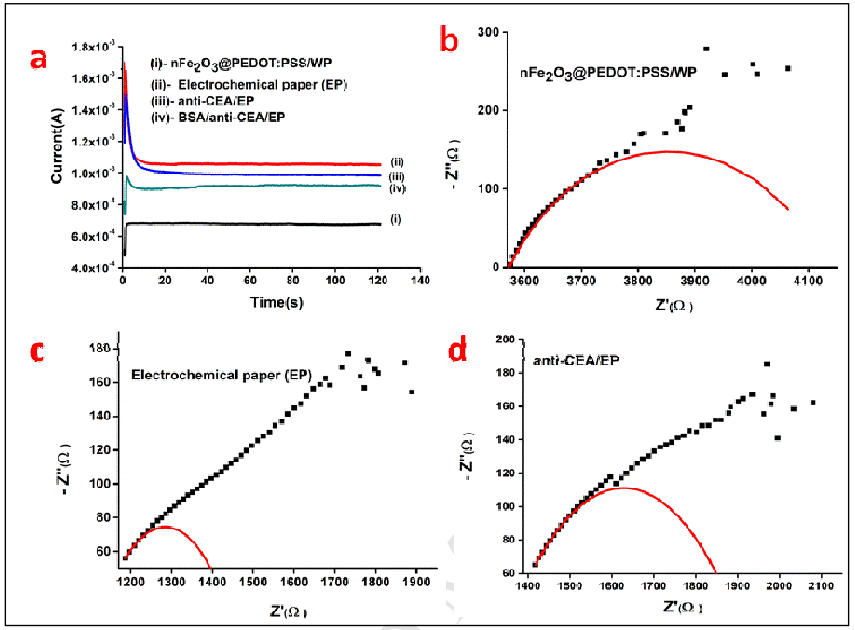


Figure 5: (a) Chronoamperometric studies of various modified paper electrode. (b-d) EIS plot of nFe_2O_3/PEDOT:PSS/WP, electrochemical paper (EP), and anti-CEA/EP.

Electrochemical impedance spectroscopy (EIS) was used to measure the impedance by applying small amplitude alternating current [34]. Quantification of the R_{ct} (charge transfer resistance) value of the EP electrode was done by analyzing semicircle arc of the Nyquist plot. Figure 5 shows the Nyquist plot obtained for (b) iron oxide nanoparticles decorated PEDOT:PSS/WP (nFe_2O_3/PEDOT:PSS/WP), (c) DMSO treated nFe_2O_3/PEDOT:PSS/WP, EP electrode and (d) anti-CEA/EP in 5 mM $[Fe(CN)_6]^{3-/4-}$ containing PBS solution at the biasing potential of 0.03 V and the frequency range from 50 Hz to 100 KHz. It was observed that R_{ct} value of nFe_2O_3/PEDOT:PSS/WP (966.39 Ω, Figure 5b) was found to be decreased on its treatment with high dielectric solvent DMSO (368.66 Ω, figure 5c) which diminished the

coulomobic interaction between PEDOT and PSS resulting in ejection of the insulating PSS from the surface of conducting polymer [35]. This PSS ejection enhanced the exposure of inter-connected PEDOT chains resulting in enhanced $[Fe(CN)_6]^{3-/4-}$ redox probe permeability to the surface of EP electrode. After immobilization of anti-CEA onto EP electrode (anti-CEA/EP), the R_{ct} value increased to 632.08 Ω (figure 5d) due to presence of insulating surface for charge transfer [36]. Heterogeneous electron transfer rate constant (K_{ct}) of the fabricated electrodes was idetermined to delineate the electron transfer kinetics by using Eq. (1) [37]:

$$K_{ct} = RT/(n^2 F^2 A R_{ct} [S]) \qquad \text{Eq. (1)}$$

where R is ideal gas constant, T is room temperature (298 K), F is Faraday constant (96485 J), A is electroactive surface area (1 cm^2), [S] is the concentration of redox probe (mol/cm^3) and n is the number of the electrons transferred per molecule of the redox probe. By using Eq. 1, K_{ct} value of the nFe_2O_3/PEDOT:PSS/WP and EP was found to be 5.51 $x10^{-4}$ and 14.45 $x10^{-5}$, respectively. Thus, EP electrode showed increased electron transfer kinetics. The observed electrochemical properties and the conductivity of EP electrode are shown in the Table 2.

Table 2: Electrochemical properties of modified paper electrode.

S. No.	Electrode Surface	Conductivity (Scm^{-1})	R_{ct}	R_s	K_{ct}
1	PEDOT:PSS on Whatman filter paper [Conducting Paper (CP)]	6.84 x 10^{-4}	2.30 kΩ	2.16 kΩ	2.32 x 10^{-5}
2	nFe_2O_3/CP	9.3 x 10^{-3}	966.39Ω	2.12kΩ	5.51 x 10^{-5}
3	nFe_2O_3/DMSO@CP	2.30 x 10^{-2}	368.66Ω	1.10kΩ	14.45 x 10^{-5}

3.7 Electrochemical response studies

The electrochemical response of BSA/anti-CEA/EP electrode was recorded (figure 6a) as a function of increasing concentration of CEA (2-100 $ngmL^{-1}$) in 5mM $[Fe(CN)_6]^{-3/-4}$ containing PBS (50 mM, pH 7.0) solution using chronoamperometry with an incubation period of 10 minutes. The response current was found to be decreased with increase in the concentration of CEA due to the formation of electrically insulating layer of antigen-antibody complex onto the surface of fabricated biosensing electrode that perhaps hindered the electron transfer *via* the redox probe, $[Fe(CN)_6]^{3-/4-}$ at the surface of EP electrode [36]. Figure 6(b) shows the calibration curve obtained between the CEA concentration and electrochemical response current. It was observed that the electrochemical current decreased from 2 to 25 $ngmL^{-1}$ after which the saturation in current response was observed up to 100 ng mL^{-1}. A linear detection range was obtained for CEA concentration in between 4-25 $ngmL^{-1}$ with sensitivity of 10.2 $\mu Ang^{-1} mLcm^{-2}$. The linear calibration curve obeyed Eq 2.

$$I(A) = -10.2\ \mu A\ mL\ ng^{-1} \times [\text{CEA concentration}] + 746\ \mu A;\ R^2 = 0.95 \qquad \text{Eq. 2}$$

The EP electrode was fabricated without nFe_2O_3 doping (with similar procedure as discussed in Section 2.2 and 2.3) and electrochemical response studies were conducted (please see supporting information, figure S1, a) as a function of CEA concentration (2-100 $ngmL^{-1}$). Figure S1,b shows the calibration curve obtained between the CEA concentration and electrochemical response current. A linear detection range was obtained for CEA concentration in between 4-50 $ngmL^{-1}$ with sensitivity of 2.1 $\mu Ang^{-1}mLcm^{-2}$ which is ~4.8 times lower than EP electrode performance. This confirmed that nFe_2O_3 modified EP sensor provided improved conduction of electronic and ionic charge carriers within the hierarchical structure of cellulose paper resulting in improved sensitivity.

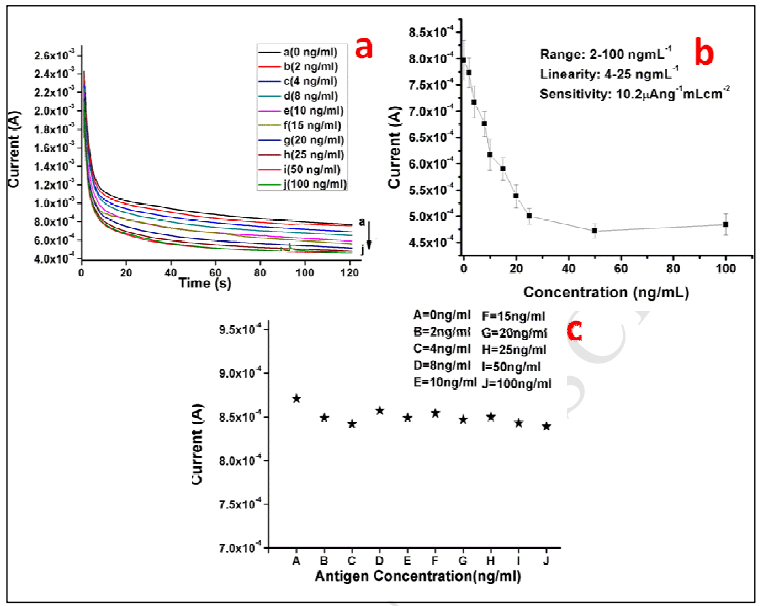


Figure 6: (a) Electrochemical response studies of BSA/anti-CEA/EP electrode obtained as a function of CEA concentration (2-100 $ngmL^{-1}$) using chronoamperometry **(b)** Calibration plot between the magnitudes of current recorded and CEA concentration **(c)** Control study plot.

To investigate the cross reactivity of the fabricated electrode (BSA/EP), the control studies were carried out. In this experiment, an electrochemical current response was measured against the increasing concentration of CEA antigen (figure 6c) for an electrode on which antibodies were not immobilized. No significant deviation in electrochemical current response was observed.

3.8 Selectivity, reproducibility and stability studies

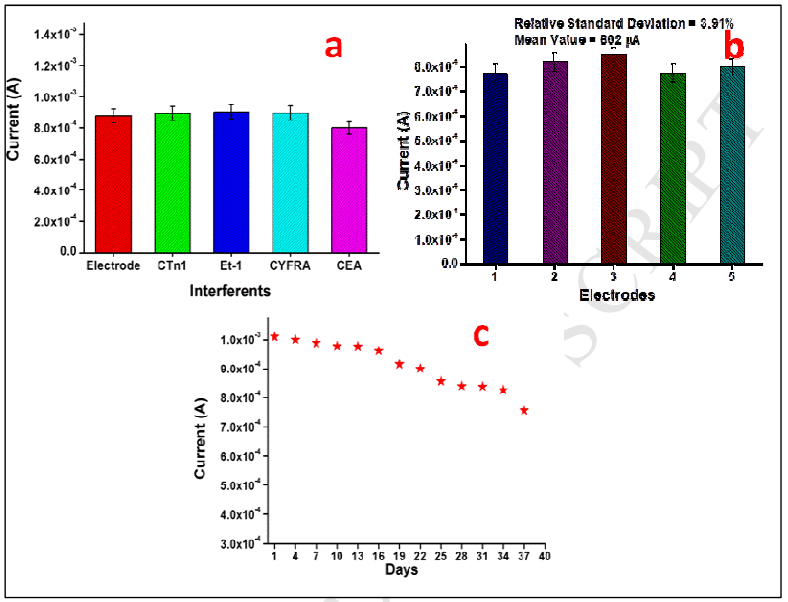


Figure 7: (a) Electrochemical current response of modified electrochemical paper electrode (BSA/anti-CEA/EP, immunoelectrode) in the presence of other analytes. **(b)** Electrochemical current response of different immunoelectrode fabricated via same set of procedure **(c)** Electrochemical current response of anti-CEA immobilized immunoelectrode measured as a function of time (day).

The selectivity of the fabricated EP bioelectrode (BSA/anti-CEA/EP) was examined in the presence of cardiac troponin I (cTnI), endotheline-1 (ET-1) and cytokeratin-19 fragment (CYFRA) concentration present in normal human serum are 1 $ngmL^{-1}$, 2-4 $pgmL^{-1}$ and 3.5 $ngmL^{-1}$ [38-40]. We performed interference studies even at concentration of 2 $ngmL^{-1}$ and observed (figure 7a) no significant change in electrochemical current. And we did not find any significant change in amperometric current on addition of higher concentration of CYFRA (4 $ngmL^{-1}$). For the reproducibility study, the electrochemical current response of five different

electrodes (BSA/anti-CEA/EP) having same surface area, fabricated under a similar set of conditions was noted in the presence of CEA concentration (2 ngmL^{-1}) [Figure 7(b)]. This electrochemical paper electrode was found to be reproducible after the relative standard deviation of less than 5 % (Mean value = 802 µA) was observed. Further, the experiments were conducted 3-times for each electrode and the error bars were included accordingly. Further, the storage stability experiment on conducting paper biosensor electrode (BSA/anti-CEA/EP) was conducted at an interval of three days (Figure 7 (c)). It was observed that 90% activity remained upto 19 days and after 34 days the EP biosensor electrode retained the activity up to 82% when it was stored at 4 °C after which the electrochemical response current decreased to 75% in about 37 days. It can be inferred that the fabricated electrode showed reasonable good stability up to 34 days.

The characteristics of the EP based CEA biosensor were compared with those reported in literature (Table 3). This comprative study clearly revealed that Fe_2O_3/PEDOT:PSS based conducting paper (EP) showed promising detection range, sensitivity and improved shelf life as compared to other biosensors for CEA detection.

Table 3: Comparison of present work with other electrochemical sensors reported in literature for CEA detection.

Immobilization Matrix	Fabrication Method	Detection Method	Linear Detection Range (ngmL^{-1})	Sensitivity (µAng^{-1} mLcm^{2})	Shelf life (days)	Ref.
MoS_2-PBNCs	Drop casting of materials over glassy carbon electrode	Differential Pulse Voltammetry	0.005 to 10 (in log scale)	-	14	[9]
PDA/RGO	Drop casting of materials over glassy carbon electrode	Amperometry	0.0005 to 5 (in log scale)	-	30	[10]
PEDOT/IL/PANI	Drop casting of materials over glassy carbon electrode	Differential Pulse Voltammetry	0.001 to 10 (in log scale)	-	-	[11]
RGO/PEDOT:PSS	Solution processed (filter paper dip coated & treated with ethylene glycol)	Amperometry	2-8	25.8	21	[2]
CNT/PEDOT:PSS	Solution processed (filter paper dip coated & treated with formic acid)	Amperometry	2-15	7.8	18	[3]
PANI/Au	Deposition of Au on filter paper via sputtering and electro-chemical deposition of PANI	Amperometry	2-20	13.9	22	[4]
PEDOT:PSS-PVA electrospun	Electrospinning of materials on conducting paper	Amperometry	0.2-25	14.2	22	[5]
Fe_2O_3/PEDOT:PSS	Solution processed (filter paper dip coated & treated with DMSO)	Amperometry	4-25	10.2	34	Present work

MoS_2: Molybdenum disulfide, PBNCs: Prussian blue nanocubes, PDA: polydopamine, RGO: Reduced graphene oxide, IL: Ionic liquid, PANI: Polyaniline, PEDOT: Poly(3,4-ethylenedioxythiophene), PSS: Polystyrene sulfonate, CNT: Carbon nanotube, Au: Gold, PVA: Polyvinyl alcohol, Fe_2O_3: iron oxide nanoparticles

3.9 Real sample analysis

Real sample analysis was conducted to investigate the feasibility of the fabricated EP biosensor electrode (BSA/anti-CEA/EP). The blood samples of cancer patients were collected and pre-treated at RGCI&RC (Rajiv Gandhi Cancer Institute & Research Centre), Rohini, Delhi, India. Prior to conducting the real sample analysis experiments, we obtained the ethical approval from Institutional Ethical and Biosafety Committee, DTU and Institutional Review Board of RGCI&RC. The concentration of CEA present in the serum sample of cancer patients was measured using the VITROS Immunodiagnostic System. The sensitivity of EP sensor in standard and serum sample of CEA was found to be ~ 10.6 µAng^{-1}mLcm^{-2} and 14 µAng^{-1}mLcm^{-2} in the range of 4-20 ngmL^{-1} (figure S2). The sllight deviation in sensitivity of serum samples was perhaps due to presence of other compounds present in serum samples of a person. A reasonable

correlation was found between (a) CEA concentration in serum samples of cancer patients and (b) CEA concentration in standard samples as shown in Table 4.

Table 4: Determination of CEA concentration in serum samples by BSA/anti-CEA/EP electrode

CEA Concentration ($ngmL^{-1}$)	Value of current obtained with serum sample (mA)	Value of current obtained for standard sample (mA)	% RSD
2	0.753	0.773	1.9
4	0.711	0.717	0.6
8	0.648	0.676	3
10	0.568	0.617	5.8
15	0.527	0.59	8
20	0.473	0.538	9

4. CONCLUSION

Nanohybrid of nFe_2O_3@PEDOT:PSS has been used to obtain a flexible, lightweight and disposable paper based electrodes. The conductivity and electrochemical studies have revealed that this electrode exhibits excellent electrical conductivity and high heterogeneous electron transfer rate constant. The nanostructured paper based electrodes has been utilized for the immobilization of anti-CEA monoclonal antibodies. The resulting immunoelectrode (BSA/anti-CEA/nFe_2O_3@PEDOT:PSS/WP) has been used for detection of the cancer biomarker (CEA). The proposed electrode has been found to be highly sensitive (10.2 $\mu Ang^{-1} mL\ cm^{-2}$) with good linearity in the physiological range 4-25 ng mL^{-1}. Further, a good storage stability (34 days) and applicability in real patient samples indicates the potential of this lightweight, disposable and sensitive paper based platform for smart point-of-care devices.

ACKNOWLEDGEMENTS

The authors thank Prof. Yogesh Singh, Vice Chancellor, Delhi Technological University (DTU), Delhi, India for providing the facilities. Saurabh Kumar is thankful to Prof. Shyam S. Pandey, Kyushu Institute of Technology, Japan for XPS studies and Chandra Mouli Pandey, DTU for providing iron oxide nanoparticles. S.K. is thankful to Dr. Birendra K. Yadav, Dr. D. C. Doval, Rajiv Gandhi Cancer Institute & Research Centre for providing the clinical samples. Saurabh Kumar acknowledges CeNSE, Indian institute of science, Bangalore, India and Department of Science and Technology, New Delhi, India for DSTINSPIRE Faculty Award (DST/INSPIRE/04/2017/002750). Suveen Kumar acknowledges the Department of Science and Technology, New Delhi, India for the DST-INSPIRE Faculty Award (DST/INSPIRE/04/2017/001336).

REFRENCES

[1] E. Morgan, I. Wilson, An early description of paper chromatography?, Chromatographia, 60 (2004) 135-136.
[2] D.D. Liana, B. Raguse, J.J. Gooding, E. Chow, Recent advances in paper-based sensors, Sensors, 12 (2012) 11505-11526.
[3] B. Liu, D. Du, X. Hua, X.Y. Yu, Y. Lin, Paper-based electrochemical biosensors: from test strips to paper-based microfluidics, Electroanalysis, 26 (2014) 1214-1223.
[4] A.W. Martinez, S.T. Phillips, M.J. Butte, G.M. Whitesides, Patterned paper as a platform for inexpensive, low-volume, portable bioassays, Angewandte Chemie International Edition, 46 (2007) 1318-1320.
[5] S. Gong, W. Schwalb, Y. Wang, Y. Chen, Y. Tang, J. Si, B. Shirinzadeh, W. Cheng, A wearable and highly sensitive pressure sensor with ultrathin gold nanowires, Nature Communications, 5 (2014) 3132.
[6] A. Manekkathodi, M.Y. Lu, C.W. Wang, L.J. Chen, Direct growth of aligned zinc oxide nanorods on paper substrates for low-cost flexible electronics, Advanced Materials, 22 (2010) 4059-4063.
[7] W.J. Hyun, O.O. Park, B.D. Chin, Foldable graphene electronic circuits based on paper substrates, Advanced materials, 25 (2013) 4729-4734.
[8] L. Wang, W. Chen, D. Xu, B.S. Shim, Y. Zhu, F. Sun, L. Liu, C. Peng, Z. Jin, C. Xu, Simple, rapid, sensitive, and versatile SWNT− paper sensor for environmental toxin detection competitive with ELISA, Nano letters, 9 (2009) 4147-4152.
[9] S. Kumar, A. Sen, S. Kumar, S. Augustine, B.K. Yadav, S. Mishra, B.D. Malhotra, Polyaniline modified flexible conducting paper for cancer detection, Applied Physics Letters, 108 (2016) 203702.

[10] S. Kumar, P. Rai, J.G. Sharma, A. Sharma, B.D. Malhotra, PEDOT: PSS/PVA-nanofibers-decorated conducting paper for cancer diagnostics, Advanced Materials Technologies, 1 (2016) 1600056.
[11] S. Kumar, S. Kumar, S. Srivastava, B.K. Yadav, S.H. Lee, J.G. Sharma, D.C. Doval, B.D. Malhotra, Reduced graphene oxide modified smart conducting paper for cancer biosensor, Biosensors and Bioelectronics, 73 (2015) 114-122.
[12] D. Bloor, B. Movaghar, Conducting polymers, IEE Proceedings I-Solid-State and Electron Devices, 130 (1983) 225-232.
[13] M. Gerard, A. Chaubey, B. Malhotra, Application of conducting polymers to biosensors, Biosensors and Bioelectronics, 17 (2002) 345-359.
[14] H. Okuzaki, M. Ishihara, S. Ashizawa, Characteristics of conducting polymer transistors prepared by line patterning, Synthetic metals, 137 (2003) 947-948.
[15] S. Ashizawa, Y. Shinohara, H. Shindo, Y. Watanabe, H. Okuzaki, Polymer FET with a conducting channel, Synthetic metals, 153 (2005) 41-44.
[16] J. Jang, M. Chang, H. Yoon, Chemical sensors based on highly conductive poly (3, 4-ethylenedioxythiophene) nanorods, Advanced Materials, 17 (2005) 1616-1620.
[17] A.L. Briseno, M. Roberts, M.-M. Ling, H. Moon, E.J. Nemanick, Z. Bao, Patterning organic semiconductors using “dry” poly (dimethylsiloxane) elastomeric stamps for thin film transistors, Journal of the American Chemical Society, 128 (2006) 3880-3881.
[18] H. Yan, T. Kagata, Y. Mori, Y. Harashina, Y. Hara, H. Okuzaki, Micrometer-scaled channel lengths of organic field-effect transistors patterned by using PEDOT/PSS microfibers, Chemistry Letters, 37 (2008) 44-45.
[19] P.R. Solanki, A. Kaushik, V.V. Agrawal, B.D. Malhotra, Nanostructured metal oxide-based biosensors, NPG Asia Materials, 3 (2011) 17-24.
[20] B. Stewart, C.P. Wild, World cancer report 2014, World, (2016).
[21] A.N. Bhatt, R. Mathur, A. Farooque, A. Verma, B. Dwarakanath, Cancer biomarkers-current perspectives, 132 (2010) 129-149.
[22] J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, M. Rebelo, D. Parkin, D. Forman, F. Bray, GLOBOCAN 2012 v1.0, cancer incidence and mortality worldwide: IARC cancer base no. 11 [online]. International Agency for Research on Cancer, Lyon (France) 2013.
[23] L. Zhang, Y. Wang, L. Shen, J. Yu, S. Ge, M. Yan, Electrochemiluminescence behavior of AgNCs and its application in immunosensors based on PANI/PPy-Ag dendrite-modified electrode, Analyst, 142 (2017) 2587-2594.
[24] G. Nie, Y. Wang, Y. Tang, D. Zhao, Q. Guo, A graphene quantum dots based electrochemiluminescence immunosensor for carcinoembryonic antigen detection using poly (5-formylindole)/reduced graphene oxide nanocomposite, Biosensors and Bioelectronics, 101 (2018) 123-128.
[25] G. Nie, Y. Tang, B. Zhang, Y. Wang, Q. Guo, Label-free photoelectrochemical immunosensing platform for detection of carcinoembryonic antigen through photoactive conducting poly (5-formylindole) nanocomposite, Biosensors and Bioelectronics, 116 (2018) 60-66.
[26] D. Zhao, Y. Wang, G. Nie, Electrochemical immunosensor for the carcinoembryonic antigen based on a nanocomposite consisting of reduced graphene oxide, gold nanoparticles and poly (indole-6-carboxylic acid), Microchimica Acta, 183 (2016) 2925-2932.
[27] C.J. Jeong, S.M. Sharker, I. In, S.Y. Park, Iron oxide@ PEDOT-based recyclable photothermal nanoparticles with poly (vinylpyrrolidone) sulfobetaines for rapid and effective antibacterial activity, ACS Applied Materials & Interfaces, 7 (2015) 9469-9478.
[28] K. Singh, A. Ohlan, P. Saini, S. Dhawan, Poly (3, 4-ethylenedioxythiophene) γ-Fe2O3 polymer composite–super paramagnetic behavior and variable range hopping 1D conduction mechanism–synthesis and characterization, Polymers for Advanced Technologies, 19 (2008) 229-236.

[29] K. Aasmundtveit, E. Samuelsen, O. Inganäs, L. Pettersson, T. Johansson, S. Ferrer, Structural aspects of electrochemical doping and dedoping of poly (3, 4-ethylenedioxythiophene), Synthetic Metals, 113 (2000) 93-97.
[30] L. Ouyang, C. Musumeci, M.J. Jafari, T. Ederth, O. Inganäs, Imaging the phase separation between PEDOT and polyelectrolytes during processing of highly conductive PEDOT: PSS films, ACS applied materials & interfaces, 7 (2015) 19764-19773.
[31] T.-R. Chou, S.-H. Chen, Y.-T. Chiang, Y.-T. Lin, C.-Y. Chao, Highly conductive PEDOT: PSS films by post-treatment with dimethyl sulfoxide for ITO-free liquid crystal display, Journal of Materials Chemistry C, 3 (2015) 3760-3766.
[32] D.A. Mengistie, M.A. Ibrahem, P.-C. Wang, C.-W. Chu, Highly conductive PEDOT: PSS treated with formic acid for ITO-free polymer solar cells, ACS Applied Materials & Interfaces, 6 (2014) 2292-2299.
[33] Y. Xia, K. Sun, J. Ouyang, Solution-processed metallic conducting polymer films as transparent electrode of optoelectronic devices, Advanced Materials, 24 (2012) 2436-2440.
[34] M. Moisel, M.L. de Mele, W.D. Müller, Biomaterial interface investigated by electrochemical impedance spectroscopy, Advanced Engineering Materials, 10 (2008) B33-B46.
[35] S. Liu, H. Deng, Y. Zhao, S. Ren, Q. Fu, The optimization of thermoelectric properties in a PEDOT: PSS thin film through post-treatment, RSC Advances, 5 (2015) 1910-1917.
[36] S. Kumar, M. Willander, J.G. Sharma, B.D. Malhotra, A solution processed carbon nanotube modified conducting paper sensor for cancer detection, Journal of Materials Chemistry B, 3 (2015) 9305-9314.
[37] A. Bardea, E. Katz, I. Willner, Probing antigen–antibody interactions on electrode supports by the biocatalyzed precipitation of an insoluble product, Electroanalysis: An International Journal Devoted to Fundamental and Practical Aspects of Electroanalysis, 12 (2000) 1097-1106.
[38] D. Sandil, S. Kumar, K. Arora, S. Srivastava, B. Malhotra, S. Sharma, N.K. Puri, Biofunctionalized nanostructured tungsten trioxide based sensor for cardiac biomarker detection, Materials Letters, 186 (2017) 202-205.
[39] E. Asham, A. Shankar, M. Loizidou, S. Fredericks, K. Miller, P. Boulos, G. Burnstock, I. Taylor, Increased endothelin-1 in colorectal cancer and reduction of tumour growth by ET A receptor antagonism, British Journal of Cancer, 85 (2001) 1759.
[40] K. Okamura, K. Takayama, M. Izumi, T. Harada, K. Furuyama, Y. Nakanishi, Diagnostic value of CEA and CYFRA 21-1 tumor markers in primary lung cancer, Lung Cancer, 80 (2013) 45-49.

ACCEPTED MANUSCRIPT

Highlights

1. PEDOT:PSS and nanostructured iron oxide modified conducting paper sensor has been reported for cancer biomarker (CEA) detection.
2. The electrical conductivity of the conducting paper electrode has been enhanced by two orders of magnitude after treatment with dimethyl sulfoxide.
3. This low cost, flexible, and disposable paper electrode exhibits high sensitivity of 10.2 $\mu A\ ng^{-1}mL\ cm^{-2}$ in the detection range of 4-25 $ngmL^{-1}$ for CEA detection.
4. It may be a promising alternative over the expensive conventional electrodes (ITO, gold and glassy carbon).

ACCEPTED MANUSCRIPT

IJASRM

International Journal of Advanced Scientific Research and Management, Volume 3 Issue 12, Dec 2018

www.ijasrm.com

ISSN 2455-6378

Empowerment of Women in India

Neha Tomar[1] and Seema Singh[2]

[1]Post Doctoral Fellow,Department of Humanities,Delhi Technological University, Delhi, India

[2]Professor of Economics,Department of Humanities
Delhi Technological University, Delhi

Abstract

The Problem – Why we all talk about women empowerment and not men? The subject of empowerment of women has becoming a burning issue all over the world including India since last few decades. Women constitute almost 50% of the world's population but India has shown disproportionate sex ratio whereby female's population has been comparatively lower than males. As far as their social status is concerned, they are not treated as equal to men in all the places in India. Empowerment of women means developing them as more aware individuals, who are politically active, economically productive and independent and are able to make intelligent discussion in matters that affect them **The Solution** – The purpose of this paper is to throw light on historical background of women empowerment in India. The paper addresses three questions: What is the evidence base to support that investing in women is beneficial ? What are the major constraints on realizing the full potential of women in the process of economic development? How these constraints can be unblocked? **The Methodology –** The research is based on the literature collected from journals of international repute such as Oxford University Press, Sage Publications, Emerald, Taylor and Francis etc. and Working papers of World Bank, United Nation, Organization for Economic Co-operation and Development, Asian Development Bank and International Labor Organization and International Monetary Fund **Research limitations/implications** –. The study will have important implications for further designing of new policies and giving new insights to the role of women in the economy. **Keywords** Women empowerment, social issues, Economic Development, Economic Growth. Women and Economy.

.1. Introduction

Mahatma Gandhi stated that the position of women the society is an index of its civilization. "Train a man and you train an individual, Train a woman and you build a nation". Women are an integral part of every economy. All round development and harmonious growth of a nation would be possible only when women are considered as equal partners in progress with men. . Empowerment can be viewed as means of creating a social environment in which one can make decisions and make choices either individually or collectively for social transformation. Although women constitute half of world's population yet they are the largest group which is excluded from the benefits of social and economic development. Women constitute a strong labor force which needs to be mobilized and encouraged to make an effective contribution to the development process. In India, the work participation rate for women is less than half that of men. Empowerment of women has become a subject of growing importance around the world in contemporary times. The concern is seen at different levels and circles which include government, bureaucracy, nongovernmental organizations, researchers and women's empowerment. There has also been a shift in the demands of women from equality in 1970s to development in the 1980s to empowerment- social, political and economic since the 1990s. Since the term empowerment came in to be used from 80s.

Empowerment of women has came to be accepted as a key objective of development. Empowerment is a process of improving women's control over resources, ideology and self. It is a process of enlarging the entitlements and choices of women.

Although India has achieved rapid economic growth during the last two decades but still gender disparities remain deep and persistent in India (Duflo 2012, World Bank 2012). Many studies both theoretical and empirical states that women are of crucial importance to process of introducing innovations into markets, they create employment (including self-employment) contributing to overall wealth creation in all economies (Brush 2006). An increase in female employment could significantly boost growth and per capita income (ILO -2015 and ILO -2014) Advancing gender equality requires strengthening different dimensions of women's autonomy: economic and political autonomy, full citizenship and freedom from all forms of violence, and sexual and reproductive autonomy (AlpízarDurán, 2010).Greater gender equality in employment and education emerge as important determinants of growth but it also appears pivotal to the translation of growth into progress on gender equality across a broad range of fronts (Kabeer and Natali, 2013). Gender inequality in education has a significant negative impact on economic growth and addition to increasing growth, greater gender equality in education promotes other important development goals(Stephan Klasen, 1999).

2. Conceptualizing Women Empowerment

"There is no tool for development more effective than the empowerment of women." Kofi Annan. According to (NaliaKabeer's 2001)to be disempowered means to be denied choice, while empowerment refers to the processes by which those who have been denied the ability to make choices acquire such an ability. According to Kabeer, the concept of empowerment can be explained through three closely interrelated dimensions: agency, resources, and achievements. Agency refers to the processes by which choices are made and put into effect and is central to the concept of empowerment whereas resources are the medium through which agency is exercised and achievements implies outcomes of the agency. How women exercise choice and the actual outcomes will depend on the individual. Choices will vary across class, time and space. Moreover, impacts on empowerment perceived by outsiders might not necessarily be those most valued by women themselves. McWhirter (1991) defines empowerment as: The process by which people, organizations or groups who are powerless (a) become aware of the power dynamics at work in their life context, (b) develop the skills and capacity for gaining some reasonable control over their lives, (c) exercise this control without infringing upon the rights of others and (d) support the empowerment of others in the community. Empowerment is about working to remove the existence and effects of unjust inequalities (Ward and Mullender, 1991). Empowerment can take place on a small scale, linking people with others in similar situations through self-help, education, support, or social action groups and network building; or on a larger scale, through community organization, campaigning, legislative lobbying, social planning, and policy development (Parsons, 1991). Empowerment is a process whereby women become able to organize themselves to increase their own self-reliance, to assert their independent right to make choices and to control resources which will assist in challenging and eliminating their own subordination(Keller and Mbwewe, 1991).

Empowerment is basically autonomy at both collective and individual level. It encompasses several mutually reinforcing components but begins with and is supported by economic independence(Hapke,1992). Empowerment is the process in which people assume an increased involvement in defining and promoting their own agenda for political and social projects designed to enact change (Wilkinson,1990). It is a process aimed at changing the nature and direction of systemic forces which marginalize women and other disadvantaged sections in a given context (Sharma,1991and 92). Empowerment is the vehicle which enables women to renegotiate their existence on an equitable basis (Banerjee,1992). Empowerment is a process to participate effectively, in decisions that affect women's lives at the family, community and higher levels of the political process (ISED, 1991). The empowerment process is facilitated by creating awareness about one's rights and responsibilities and socioeconomic, educational and political opportunities, by developing skills for utilizing productive resources and by involving oneself in collective activities and community life (Pandey,1993).Empowerment requires transformation of structures of subordination through radical changes in law, property rights, and other institutions that reinforce and perpetuate male domination (Sen and Grown,1985). Empowerment is all about empowering the powerless, 'power ' is fundamental to the enabling process of empowerment. Power has been defined as "the ability to accomplish one's will regardless of the means employed" (March and Taqque, 1982). Women's empowerment is a multi-dimensionalprocess of civil, political, social, economic, andcultural participation and rights. It is the ''frame'' of the gender indicators which reflects that an enhanced participation and rights inthese domains underlie women's equal citizenship(Moghadam and Senftovan, 2005). The concept of 'empowerment', if it is used precisely and deliberately, can help to focus thought, planning, and action in development. However, when its use is careless, deliberately vague, or sloganizing, it risks becoming degraded and valueless. (Rowlands,1995) Thus empowerment and autonomy of women refers to improvement intheir political, social, economic and health status and overall development.

3.Status of Women in India

The status of Women in India has been subject to many great changes over the past few millennia. Women during the early Vedic period enjoyed equal status with men in all aspects of life. Works by ancient Indian grammarians such as Patanjali and Katyayana suggest that women were educated in the early Vedic period. Rigvedic verses suggest that women married at a mature age and were probably free to select their own husbands in a practice called swayamvar or live-in relationship called Gandharva marriage. The position of Indian women in society further deteriorated during the medieval period when child marriages and a ban on remarriage by widows became part of social life in some communities in India. The Muslim conquest in the Indian subcontinent brought purdah to Indian society. Among the Rajputs of Rajasthan, the Jauhar was practised. In some parts of India, some of Devadasis were sexually exploited. Polygamy was practised among Hindu Kshatriya rulers for some political reasons. In many Muslim families, women were restricted to Zenana areas of the house. During the British Raj, many reformers such as Ram Mohan Roy, Ishwar Chandra Vidyasagar and Jyotirao Phule fought for the betterment of women. With a decline in their status from the ancient to medieval times,[to the promotion of equal rights by many reformers, their history has been eventful. When our country got its independence, the participation of women nationalists was widely acknowledged. When the Indian Constitution was formulated, it granted equal rights to women, considering them legal citizens of the country and as an equal to men in terms of freedom and opportunity. The sex ratio of women at this time was slightly better than what it is today, standing at 945 females per 1000 males. Yet the condition of women screamed a different reality.They were relegated to the household, and made to submit to the male-dominated patriarchal society, as has always been prevalent in our country. Indian women, who fought as equals with men in the nationalist struggle, were not given that free public space anymore. They became homemakers, and were mainly meant to build a strong home to support their men who were to build the newly independent country. Women were reduced to being second class citizens. In 1951, the first Census taken four years after independence, the picture was not much better. The female literacy rate stood at 7.93 per cent, as compared to 24.95 per cent for men. The 2001 Census suggests a 65.38 per cent literacy rate for India, with 75.85 per cent for males, and 54.16 per cent for females. The number of women goes down as we move to higher levels of education. Females constitute about 50% of country's human resource but lack of education snatches their chance to be a part of the progress and development of India. This means our pace of progress is less than the required pace. Even if females do not use education to work, total illiteracy has a huge negative impact on our society.

4. History of Women Empowerment in India

Women empowerment as a concept was introduced at the International women Conference in 1985 at Nairobi, which defined it as redistribution of social power and control of resources in favour of women.(U. Koko, 1992). Batliwala (1974) defines empowerment as "the process of challenging existing power relation and of gaining greater control over the source of power).

In the process of poverty eradication and reducing gender discrimination, the governments have been implementing various schemes and programmes providing ways and means towards women development and empowerment. The constitution of India has given special attention to the needs of women to enable them to exercise their rights on equal footing with men and participate in national development Framing of the five year plans was the first major step taken in the direction of welfare state: Jawaharlal Nehru, the first Prime Minister of India and the pioneer of five year plans, stressed on welfare of women, children and tribals in our countryThe position and status of women all over the world has risen incredibly in the 20th century. We find that it has been very low in 18th and 19th centuries in India and elsewhere when

they were treated like 'objects' that can be bought and sold. For a long time women in India remained within the four walls of their household. Their dependence on men folk was total.

A long struggle going back over a century has brought women the property rights, an equality in civil rights before the law in matters of marriage and employment

There was a time when women's education was not a priority even among the elite. Since the last quarter of the 20th century and more so after the opening up of die economy, post-1991, a growing number of women have been entering into the economic field, seeking paid work (remunerative jobs) outside the family. Women are playing bigger and bigger role in economic field: as workers, consumers, entrepreneurs, managers and investors.

The Ministry of Women and Child Development (MWCD) and the Central Social Welfare Board (CSWB) are implementing various welfare schemes and programmes for Indian women. The constitution of India provides a number of rights to women and several laws exist to protect women from discrimination. But still Indian women cannot say with pride that they possess complete empowerment equal to men. Still a lot has to be done to make them fully self-dependent and empowered equally with men. Problems like female feticide, torture for dowry, violence against women etc. still restrict the complete empowerment of women (Kapur, 2001).

Prof. Dastgir Alam, (Alam, 2008) in his paper, stated that women have always got a lesser share in labour force participation rate (LFPR). He has described the concept of social justice as a sense of participation in all spheres of human activities, social, economic and political, in the mind of masses. The article written by Smt. Sushila Kaushik (2002) throws light on the different issues responsible for women's achievement as well as the pitfalls. She states that, even if the funds and credit is available to women, one should see whether it is available to women within the households. The author states further that, to improve the status of women, there is a need to integrate them into national structure and policy making rather than have policies on them.

Apart from Government organizations there are other non government organizations and groups which also contributed majorly in empowerment of women. Self-help groups emerged as initiatives by people to earn their livelihood by pooling a common fund and coming together in their effort. It was the effort of the group to stand by each other, help each other and grow together. The success and potential of the self-help groups was recognized by NABARD and also the Government of India. Women members of the self-help groups have shown increased levels of socio-political awareness and empowerment in the community, raised levels of negotiating power, and changes in community norms, particularly in terms of changing attitudes to gender expectations. (Dash A., 2003).

5. Indian constitution and women empowerment

The principle of gender equality is enshrined in the Indian Constitution in its Preamble, Fundamental Rights, Fundamental Duties and Directive Principles. The Constitution not only grants equality to women, but also empowers the State to adopt measures of positive discrimination in favor of women The constitution of India has greatly enhanced the status of Indian women by throwing upon to them all its series on equal terms with men. All the men and women of India are equity entitled for individual freedom, fundamental rights including the right to participate in social, cultural, religious, educational, economic and political activities. The constitution provides for equality of sex and offers protection to women against exploitation. It has given the voting right to women and in no way treats women as second grade citizens.

Directive principles of State Policy also contains important provisions regarding women empowerment and it is the duty of the government to apply these principles while making laws or formulating any policy. Though these are not justifiable in the Court but these are essential for governance nonetheless. Some of them are:

- Article 39 (a) provides that the State to direct its policy towards securing for men and women equally the right to an adequate means of livelihood.
- Article 39 (d) mandates equal pay for equal work for both men and women.
- Article 42 provides that the State to make provision for securing just and humane conditions of work and for maternity relief.

Apart from this 73rd and 74th Constitutional Amendment of 1993, a very important political right has been given to women which is a landmark in the direction of women empowerment in India. With this amendment women were given 33.33 percent reservation in seats at different levels of elections in local governance i.e. at Panchayat, Block and Municipality elections.

Thus it can be seen that these Constitutional provisions are very empowering for women and the State is duty bound to apply these principles in taking policy decisions as well as in enacting laws.

There are also some specific laws for women in order to protect their interest and a provides a sense of security and empowerment. The list of some specific laws which were enacted by the Parliament in order to fulfill Constitutional obligation of women empowerment:

- The Equal Remuneration Act, 1976.
- The Dowry Prohibition Act, 1961.
- The Immoral Traffic (Prevention) Act, 1956.
- The Maternity Benefit Act, 1961.
- The Medical termination of Pregnancy Act, 1971.
- The Commission of Sati (Prevention) Act, 1987.
- The Prohibition of Child Marriage Act, 2006.
- The Pre-Conception & Pre-Natal Diagnostic Techniques (Regulation and Prevention of Misuse) Act, 1994.
- The Sexual Harassment of Women at Work Place (Prevention, Protection and) Act, 2013.

Government policies and schemes for women empowerment. Within the framework of a democratic polity, our laws, development policies, Plans and Programmes have aimed at women's advancement in different spheres. From the Fifth Five Year Plan (1974- 78) onwards there has been a marked shift in the approach to women's issues from welfare to development and then from Eighth Five Year Plan emphasis was shifted from development to empowerment. In recent years, the empowerment of women has been recognized as the central issue in determining the status of women. India has also ratified various International conventions and human rights instruments committing to secure equal rights of women. Key among them is the ratification of the Convention on Elimination of All Forms of Discrimination against Women (CEDAW) in1993.

In the year 2001, the Government of India launched a National Policy for Empowerment of Women. The specific objectives of the policy are as follows:

- Creation of an environment through positive economic and social policies for full development of women to enable them to realize their full potential.
- Creation of an environment for enjoyments of all human rights and fundamental freedom by women on equal basis with men in all political, economic, social, cultural and civil spheres.
- Providing equal access to participation and decision making of women in social political and economic life of the nation.
- Providing equal access to women to health care, quality education at all levels, career and vocational guidance, employment, equal remuneration, occupational health and safety, social security and public life etc.
- Strengthening legal systems aimed at elimination of all forms of discrimination against women.
- Changing societal attitudes and community practices by active participation and involvement of both men and women.
- Mainstreaming a gender perspective in the development process.
- Elimination of discrimination and all forms of violence against women and the girl child.
- Building and strengthening partnerships with civil society, particularly women's organizations.

On other hand special budget allocations are made in Five Year Plans regarding empowerment of women. Framing of the five year plans was the first major step taken in the direction of welfare state: Jawaharlal Nehru, the first Prime Minister of India and the pioneer of five year plans, stressed on welfare of women, children and tribal's in our country. The First Five Year Plan (1951–56) contemplated welfare measures for women. To implement welfare measures for the benefit of poor women, the Central Social Welfare Board (CSWB) was established to deal with the problems of women. The CSWB recognized and realized the need for organizing women into Mahila Mandals or women's club as an approach to community development.

The Second Five Year Plan (1956 – 61) intimately concentrated overall intensive agricultural

development. However, the welfare approach to women's issues was determined recognizing women as workers. Further, protection against injuries at work, maternity benefits and crèches for their children. It also suggested immediate implementation of the principal of equal pay for equal work and provision for training to enable women to compete for higher jobs.

The Third Five Year Plan (1961 – 66) sincerely recognized the greater importance of education for women which has been a major welfare strategy for women. This plan allocated the largest share for expending social welfare services and condensed courses of education. As regards to wealth, maternal and child welfare programmes were proclaimed in terms of maternal and child welfare, health education, nutrition and family planning.

Thus the emphasis on women education was continued during the Fourth Five Year Plan also (1969 – 1974). The basic policy was to promote women's welfare as the base of operation. The outlay on family planning was stepped up to reduce the birth rate through education. Immunization of pre-school children and supplemental feeding, expectant and nursing mothers8.

Need for training women in respect of income generating activities and their protection was stressed in the Fifth Five Year Plan. Further, the fifth plan also recommended a strategic programme of functional literacy to equip women with skills and knowledge to perform the functions as a good housewife. Under the health programmes, the primary objective was to provide minimum public health facilities integrated with family planning and nutrition for vulnerable groups, children, pregnant and lactating mothers The Fifth Year Plan was happened to be during the decade of International Women's decade and the submission of the Report of the Committee on the status of women in India (CSWI) "Towards Equality". The CSWI had comprehensively examined the rights and status of women in the context of changing social and economic conditions and the problems relating to the advancement of women. The CSWI reported that the dynamics of social change and development had adversely affected a large section of women and had created new imbalances and disparities

The Sixth Five Year Plan stressed the need of economic independence educational advance and access to health care and family planning as essential for women's development. So the strategy was threefold: of education, employment and health. They are independent and dependent on the total developmental process.

The Seventh Five Year Plan sought to generate awareness among women about their rights and privileges. The long term objectives of developmental programmes in the Seventh plan were to raise women's economic and social status in order to bring them into the mainstream of national development and recognized the importance of women in contributing to the various socio-economic, political and cultural activities. The seventh plan emphasized the need to open new avenues of work for women and perceive them as crucial resource for the development of the country. Another salient and crucial recognition was the need for organization of women workers and unionization. Under the plan, a new scheme, "Women's Development Corporation" has been taken up for promoting employment generating activities by supporting schemes from women's group and women from poorer sections of society. A women's development planning and monitoring cell was also set up for collection of data and monitoring of plan programmes. A very significant step therein was to identify and promote beneficiary oriented programmes which extended direct benefits to women.

The strategy in the Eighth Plan was to ensure that the benefits of development from different sectors did not bypass women and special programmes were implemented to complement the general programmes. The main objective of Eighth Plan was to extend the reach of services to women both qualitatively and quantitatively. Panchayati Raj institutions are involved in the designing and implementation of women's programmes. The approach of the Eighth Plan made a definite shift from development to empowerment of women. In order to meet the needs of women and children, there had been a progressive increase in the plan outlays over the time of eight five year plans. The outlay of Rs. 4 crores in the First Five Year Plan (1951 – 56) had gone up to Rs. 2000 Crores in the Eighth Five Year Plan18.

The Ninth Five Year Plan came into effect from April 1, 1997. An approach paper had been developed by the Planning Commission and accepted by the National Development Council, which had become basis for developing Ninth Five Year Plan. In this approach paper focus was laid on empowerment of women and people's participation in planning and implementation of strategies. Women are facing problems like feminization of poverty, inadequate investment in social sectors,

IJASRM
International Journal of Advanced Scientific Research and Management, Volume 3 Issue 12, Dec 2018
www.ijasrm.com
ISSN 2455-6378

increasing violence against women and stereotyped portrayal of women in private and state media especially television. There is necessity for information and training opportunities, reservations and social services etc., and people's involvement is necessary for the success of any programme. Empowerment is about choices and the ability exercise women's choices will be limited unless they are more involved in policy-making. The 9th Five Year Plan is an attempt to bring in women's issues within the policy-making spheres.

The Tenth Five-Year Plan (2002-07) aims at empowering women through translating the recently adopted National Policy for Empowerment of Women (2001) into action and ensuring Survival, Protection and Development of women and children through rights based approach.

The Eleventh Five Year Plan proposes to undertake special measures for gender empowerment and equity.. Encourage women's cooperatives for self-employment. The Eleventh Five Year Plan was aimed at inclusive development. Its vision was that every woman in the country should be able to develop to her full potential and share the benefits of economic growth and prosperity. Towards this end the approach adopted was to empower women and recognize their Agency thereby seeking to make them partners in their own development.

The vision for the Twelfth Five Year Plan is to ensure improving the position and condition of women by addressing structural and institutional barriers as well as strengthening gender mainstreaming. In Twelfth plan the Government's priority would be to consolidate the existing initiatives and interventions relating to women, build upon the achievements and also move beyond to respond to new challenges. The thrust areas are economic, social and political empowerment of women. Recognizing that economic independence is the key to improving the position of women within the family and in the society and the plan would need to focus on enhancing women's access to and control over resources. It will not only increase their presence in the work force but, more importantly will improve the quality of women's work and ensuring their upward mobility on the economic ladder.

6. Gender Equality ,Growth and Development Links

"It turns out that advancing equal opportunity and economic empowerment is both morally right and good economics, because discrimination, poverty and ignorance restrict growth, while investments in education, infrastructure and scientific and technological research increase it, creating more good jobs and new wealth for all of us." – Former President William J. Clinton's remarks for the Democratic National Convention on September 5, 2012. It is imperative that empowerment and autonomy of women and improvement in their social, economic and political status is essential for the achievement of sustainable development in all areas of life. To promote gender equality and the empowerment of women as effective ways to combat poverty, hunger and disease and to stimulate development that is truly sustainable(United Nations Millennium Declaration,2000).Economic development and growth and gender equality go hand-in-hand. As Sen (1987) demonstrates, in any model of economic development it is crucial to take into account that cooperation and conflict exist simultaneously in gender divisions

The societies that discriminate by gender tend to experience less rapid economic growth and poverty reduction than societies that treat males and females more equally, and that social gender disparities produce economically inefficient outcomes (World Bank 2001). The primary pathways through which gender systems affect growth are by influencing the productivity of labour and the allocative efficiency of the economy (World Bank 2002). In terms of productivity, for example, if the access of women farmers to productive inputs and human capital were on a par with men's access, total agricultural output could increase by an estimated 6 to 20 percent (World Bank 2001). Investment in the human capital, health and education, of women and girls is presented as a key way forward as witnessed by the MDGs. The logic is that 'educated, healthy women are more able to engage in productive activities, find formal sector employment, earn higher incomes and enjoy greater returns to schooling than are uneducated women…' (WBGDG 2003).. However women's participation can only be effective and meaningful when underlying gender power relations are transformed and when attention and support are given to women's specific knowledge and capacities. Table 1 shows the relationship between the broad attributes of Economic development and growth and their positive links to women empowerment..

Table 1: Studies on the Link between Women empowerment , Gender Equality and Economic Development

Author/ Study	Purpose	Variable	Relationship Economic development and women empowerment
Esther Duflo, 2011	Relationship between Women empowerment and economic development	Empowerment Development	Positively interrelated.
Dollar and Gatti ,1999	Relationship between Gender, Income, and Growth	Gender Inequality, Income, Education Growth & Health	Gender Inequality have Negative Impact
Klasen,2002	Relationship between Women in Education and Economic Development	Gender Inequality Education Economic Development	Negative
Kabeer,1991	Relationship between Gender, development, and training	Gender equity Women empowerment Development Training	Positive
Klasen,1999	Relationship between Women and Growth and Development	Gender inequality Growth Development	Negative
Klasen and Lamanna, 2009	Relationship between Women and growth	Gender Inequality Education Employment Health	Negative
Agarwal, 2002	Relationship between Women and environment Sustainability	Gender Inequality Environmental Sustainability	Negative
Cela, Dankelman& Stern,2014	Relationship between Women Empowerment ,Economic Development and Environmental Sustainability	Gender Equality Economic Development Environmental Sustainability	Positive
Jane,2000	Relationship between Gender and Development	Participatory Empowerment, Gender Development	Positive
Forsythe,Korzeniewicz,Durrant, 2000	Relationship between status of women and Economic Development	Status of Women GDI HDI GenderInequality	Positive
Susan Buckingham-Hatfield,2002	Link between WomenParticipation and development	WorkParticipation, Health,Education, Development Gender Equality	Positive
Psacharopoulos andTzannatos, 1989	Link between Women Participation and development	Agrarian economy Women Participaion Development Education	Positive
Beneria and Sen,1981	Link between Women's Work and development	Economic growth Women Participaion Development Education	Positive
Tsani, Paroussos, Fragiadakis,	Link between Female	Female Labour Force	Positive

IJASRM
International Journal of Advanced Scientific Research and Management, Volume 3 Issue 12, Dec 2018
www.ijasrm.com
ISSN 2455-6378

Charalambidis and Capros, 2012	Labour Force Participation and development	Participation and Economic Development	
Luci,2009	Link between Female Labour Force Participation and development	Women's labour market participation Economic Growth and Development	Positive
Nath, 1970	Link between Female Labour Force Participation and development	Female labour participation Economic Growth and Development	Positive
BertaEsteveVolart,2004	Link Between Gender Discrimination and Growth	Growth Development Labor Market GDP	Positive Relationship
Bradshaw, Castellino,2013	Women's role in economic development:	Education Employment Political Participation Economic Growth& Development	Positive
Dollar and Gatti ,1999	Relationship between Gender, Income, and Growth	Gender Inequality, Income, Education Growth & Health	Gender Inequality have Negative Impact
Agarwal, 2002	Relationship between Women and environment Sustainability	Gender Inequality Environmental Sustainability	Negative
Johnsson&Latham,2007	Relationship between gender equality as a prerequisite for sustainable development	Sustainability, Lifestyles Welfare and Development Reproductive and Social roles	Positive
Diebolt and Perrin,2013	Relationship between Women Empowerment and Sustained Growth	Women Empowerment Gender Equality Economic Output Skilled Labor	Positive
Mariam Dem,1993	Relationship between Women and Sustainable development	Women Participation Economic Activities And Development	Positive
Heyzer,2005	Relationship between Women and Sustainable development	Women's rights Empowerment Millennium Development	Positive and crucial
Brand,1996	Relationship between Women and development	Women's participation at International, National and Local level and Development	Positive
Brenner,2009	Relationship between Women and Sustainable development	Gender Mainstreaming Sustainable development GAD	Positive
S. Low and Sengupta,1996	Relationship between Women and development	Women Development	Positive
Susan Buckingham-Hatfield,2002	Link between Women and development	WorkParticipation, Health,Education, Development Gender Equality	Positive

7. Gender Mainstreaming In Growth And Development In India - Constraints And Suggestions

When more women work, economies grow The centrality of gender equality has also been articulated in the outcome document of the United Nations Conference on Sustainable Development, entitled "The future we want", adopted in 2012, which included recognition of the importance of gender equality and women's empowerment across the three pillars of sustainable development, economic, social and environmental, and resolve to promote gender equality and women's full participation in sustainable development policies, programmes and decision-making at all levels. There is growing evidence of the synergies between gender equality, on the one hand, and economic, social and environmental sustainability, on the other. For example, when women have greater voice and participation in public administration, public resources are more likely to be allocated towards investments in human development priorities, including child health, nutrition and access to employment (Chattopadhyay and Duflo, 2004). Further, certain aspects of gender equality, such as female education and women's share of employment, can have a positive impact on economic growth, although this impact is dependent on the nature of growth strategies, the structure of the economy, the sectoral composition of women's employment and labour market segregation, among other factors (Kabeer and Natali, 2013).The underlying causes and consequences of unsustainability and gender inequality are deeply intertwined and rooted in the dominant economic models (Fukuda-Parr, Heintz and Seguino, 2013). Inequality itself threatens economic sustainability, fuelling unrest and conflict and undermining the stability, level playing field and consumer demand on which growth is dependant (Stiglitz, 2012).

. Based on the literature some of the major constraints faced by Indian women are as follows.

- **Lack of employment opportunities**- Although India's economic growth is rising but cause of concern in recent times is that the jobs in sufficient numbers are not being created in sectors that could readily absorb women, especially for women in rural areas.
- **Lack of skills**- Indian female workers has low skill levels due to fewer opportunities and so it subsequently makes them less employable.
- **Limited jobs in the non-farm sectors**- We know that agricultural jobs are limited as well as seasonal and therefore it contributed to the withdrawal of women from the labour force.
- **Lack of Education and Awareness** – It is another major factor as in India female literacy rate is quite low and also in some areas, there is extreme lack of awareness about the government programmes and schemes favouring girl child and women, which can be utilized to become employable.
- **Restriction of personal liberty and autonomy**- This is another important factor for low FLPR rate. Due to patriarchal beliefs, women are restricted only to household activities and don't have liberty to come out of houses and do jobs.
- **Indian social customs and practices**- In India still women are forced to live within the boundaries of house and not are allowed to go out , take their life decisions .Women are not allowed to do work or study.
- **Burden of unpaid work**-Traditionally women are supposed to be meant only for household work which includes care work, cooking, cleaning etc. which consume most of their daily routine due to which they don't get time to do jobs.
- **Wage and Employment discrimination**- More women than men work in vulnerable, low-paid, or undervalued jobs(ILO,2012).In India generally female get les salary or wages in comparison to men. Also employer prefer to appoint men than women as they think that women are inefficient and non-serious as they also have household responsibilities on them.
- **Lack of adequate infrastructure**- Due to poor connectivity and transportation facilities women are reluctant to undertake work in the nearby areas. Thus improving connectivity between the villages and the satellite towns and small cities will also enhance female labour force participation.
- **Other Social Obstacles**-Rising crime against women is also decreasing the women participation due to safety reasons. Women find themselves unsafe even in urban areas and so they prefer to stay at home rather moving out for jobs.

Considering the constraints experienced by the females and their contribution in economic growth and development following policy initiatives are required to make them economically active. The initiatives are as follows-

- Government should focus on increasing female education level as well as creating awareness about governmental support schemes for women among public.
- More thrust should be given to skill development through professional and vocational training and job oriented courses.
- Equal opportunity to work and equal wages for equal work should be provided and in this regard definitely some government intervention is required at ground level.
- Government should focus more on microfinance support and self-help group centred activities in order to bring women into mainstream development process.
- Improving connectivity between the villages and towns and small cities will also improve female labour force participation
- More employment opportunities need to be created in the industries in order to absorb the female labour force.
- Government should also focus in creating some sort of support system for women workers such as crèches, maternity benefits, relaxed working hours, etc. in order to increase female labour force participation.

8. Conclusion

Women in India slowly started recognizing her true potential. She has started questioning the rules laid down for her by the society. As a result, she has started breaking barriers and earned a respectable position in the world. Today Indian women have excelled in each and every field from social work to visiting space station. There is no arena, which remains unconquered by Indian women. Whether it is politics, sports, entertainment, literature, technology everywhere, its women power all along. But there is still a large section of women who are uneducated, and married off before the age of 18. Families are required to supply a chaste daughter to the family of her future husband. Also very few women are actually employed in good-paying jobs, and hence parents don't see the point of spending money on girls' education. Statistics say that close to 245 million Indian women lack the basic capability to read and write, which is a large number. Only 13.9% women are employed in the urban sector, and 29% in the domestic and agriculture sector, where too a majority of women are exploited by the men. The sex ratio of India shows that the Indian society is still prejudiced against female, and a lot is yet to be achieved in this context. The path towards total gender empowerment is full of potholes. Over the years, women have made great strides in many areas with notable progress in reducing some gender gaps. Yet realities such as 11,332 women and girls getting trafficked every year, and increased practice of dowry, rape and sexual harassment hit hard against all the development that has taken place. Thus, if on one hand women are climbing the ladder of success, on the other hand she is mutely suffering the violence afflicted on her by her own family members. As compared to the past, women in modern times have achieved a lot but in reality they have to still travel a long way. Moving forward, closing gender gaps will require concerted efforts across a range of policy dimensions. In this respect, the most immediate concern for policy-makers should be to alleviate the constraints on women's freedom to choose whether or not to enter the labour market and the barriers they face once they are in the workplace (ILO,2017). . Key for economic growth is the promotion of women's economic rights which entails promoting a range of women's rights including their sexual and reproductive rights and rights to education, to mobility, to voice, to ownership, and to live free from violence(UNSDSN,2013). Although government has been making efforts to empower women in socio-economic and politico-cultural aspects, so that a welfare state and a prosperous nation can be built but still there is a gap in the efforts and its implementation at ground level. However this gap is been filled by NGO's and Self Help groups but still lot of rigorous work is still required for empowering the women.

References

[1] Afridi, F., Dinkelman, T., & Mahajan, K. (2016). Why are fewer married women joining the work force in India? a decomposition analysis over two decades., Retrieved From-http://ftp.iza.org/dp9722.pdf
[2] Aguirre, D., Hoteit, L., Rupp, C., & Sabbagh, K. (2012). Empowering the third billion: Women and the world of work in 2012. Booz and Company.
[3] Alpízar Durán, L. (2010). "The implementation of the Beijing Declaration and Platform for Action, the outcomes of the 23rd special session of the General

Assembly and its contribution to shaping a gender perspective towards the full realization of the MDGs" Keynote speech at High-Level Roundtable . 54th session of the UN Commission on the Status of Women, United Nations Headquarters NY, March 2010.

[4] Andres, L., Dasgupta, B., Joseph, G., Abraham, V., & Correia, M. (2017). Precarious drop: reassessing patterns of female labor force participation in India, Policy Research Working Paper 8024 Social Development Unit, World Bank

[5] Ahmad, Jamil. (2008). Gender Inequality and Women Empowerment ± A Review. In New Dimensions of Women Empowerment. Chapter 10, pp129-137. New Delhi: Deep and Deep Publications.

[6] Akhtar, S.M. Jawed. (2008). Empowerment of Women in India: Issues and Challenges. In New Dimensions of Women Empowerment. Chapter 8, pp96-113. New Delhi: Deep and Deep Publications.

[7] Anuradha. R. Vijaya and Reddy G. Lokanadha (July 2013) . Empowerment of Rural Women Through Vocational education and Training. Conflux Journal of Education ISSN 2320-9305 Volume 1, Issue 2, page no.8 Retrieved from: http://www.naspublishers.com.

[8] Azad, Chandra Prakash. (2008). Empowerment of Women in Post-Independence Era. In?Ajit Kumar Sinha(ed) New Dimensions of Women Empowerment. Chapter 7, pp85-95. New Delhi: Deep and Deep Publications.

[9] Bani Bora, Rajesh and V.K. Dixit,(2011). Development of Vocational Skills and Women Empowerment Through Gender Resource Centres : A Study of Delhi State, Journal of Community Mobilization and Sustainable Development Vol. 6(1), 071-076, January-June,

[10] Bardhan, K. and S. Klasen ,(2000).“On UNDP’s Revisions to the Gender- Related Development Index”, Journal of Human Development, Vol.1, pp.191-195,

[11] Barkat, A, “Women Empowerment: A Key to Human Development,2008 . Retrieved from http://www.goodgovernance.org

[12] Behara Srinivasa Rao. Women Empowerment & Planning Process. Retrieved From http://www.contentwriter.in/articles/others/women-empowerment.htm

[13] Bhashanjaly, Shaily. (2002). 'Need for Empowerment of Women. Social Welfare, 49(5), 21-22.

[14] Bose, Ashish. (2004). Empowerment of Women ± How and When? Economic & Political Weekly, 35(34), p.3005.

[15] Beneria, L., & Sen, G. (1981). Accumulation, reproduction, and" women's role in economic development": Boserup revisited. Signs: Journal of Women in Culture and Society, 7(2), 279-298., Retrieved From-: http://www.jstor.org/stable/3173878

[16] Bhalla, S., & Kaur, R. (2011). Labour force participation of women in India: some facts, some queries. Retrieved From http://www.lse.ac.uk/asiaresearchcentre/_files/arcwp40-bhallakaur.pdf

[17] Bradshaw, S., Castellino, J. and Diop, B. (2013). Women’s role in economic development: Overcoming the constraints Background paper for the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda

[18] Blumberg, R.L. (2005)“Women’s Economic Empowerment as the Magic Potion of Development?” Paper presented at the 100th annual meeting of the American Sociological Association, Philadelphia,

[19] Chattopadhyay, R. and E. Duflo ,(2001) “Women's Leadership and Policy Decisions: Evidence from a Nationwide Randomized Experiment in India”, Indian Institute of Management, Calcutta and Department of Economics ,MIT, and NBER

[20] Chenoy, D (2012), “Empowering Women through Skill Development : Challenges and Opportunities”, Power Point Presentation at the GEPD forum II, Institute of Social Studies Trust and Heinrich Boll Foundation, at the India Habitat Centre, 24 April 2012

[21] Chowdhury, S. (2011). Employment in India: What does the latest data show?. Economic and political weekly, 46(32), 23-26.

[22] Cuberes, D., & Teignier, M. (2012). Gender gaps in the labor market and aggregate productivity., Working Paper No. 2012017 (Sheffield, UK, University of Sheffield).

[23] Dash, Anup Strategies for Poverty Alleviation in India (2003). CYSD’s Holistic Approach to Empowerment through the Self-Help Group Model, IDS Bulletin 34(4), 133-14,.

[24] Duflo, E. (2012). Women Empowerment and Economic Development, Journal of Economic Literature 2012, 50(4), Pages-1051–1079, Retrieved From-https://economics.mit.edu/files/7417

[25] Dr. Rekha Singh. Status of Women in Indian Society. Retrieved From: https://www.bu.edu/wcp/Papers/Huma/HumaSing.htm

[26] Empowering Women through Skill Development - Challenges and Opportunities. Gender and Economic Policy Discussion (GEPD) Forum, 24 April 2012

[27] Elborgh-Woytek, K., Newiak, M., Kochhar, K., Fabrizio, S., Kpodar, K., Wingender, P., Clements, B. and Schwartz, G. (2013).Women, work and the economy: Macroeconomic gains from gender equity, IMF Staff Discussion Note SDN/13/10 (Washington, DC, IMF).

[28] Esteve-Volart, B.(2004.)“Gender Discrimination and Growth: Theory and Evidence from India.” Retrieved From-http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1127011.

[29] Fatima, A., & Sultana, H. (2009). Tracing out the U-shape relationship between female labor force participation rate and economic development for Pakistan. International Journal of Social Economics, 36(1/2), 182-198. .Retrieved From-http://www.emeraldinsight.com/10.1108/03068290910921253.

[30] Fletcher, E., Pande, R., & Moore, C. M. T. (2017). Women and Work in India: Descriptive Evidence and a Review of Potential Policies., Retrieved From-https://www.hks.harvard.edu/sites/default/files/centers/cid/files/publications/faculty-working-papers/women_work_india_cidwp339.pdf

[31] Forsythe, N., Korzeniewicz, R. P., & Durrant, V. (2000). Gender inequalities and economic growth: A longitudinal evaluation. Economic Development and Cultural Change, 48(3), 573-617.Retrieved from- http://unsdsn.org/wp-content/uploads/2014/02/130520-Women-Economic-Development-Paper-for-HLP.pdf

[32] Gaddis, I., & Klasen, S. (2014). Economic development, structural change, and women’s labor force participation. Journal of Population Economics, 27(3), 639-681..

[33] Garces-Ozanne, A., & Singh, A. (2017). Revisiting The Decline In India’s Female Labour Force Participation: The Rise Of Machines And Security Risks., Retrieved From-https://www.otago.ac.nz/economics/otago670806.pdf

[34] Goldin, C. (1994). The U-shaped female labor force function in economic development and economic history (No. w4707). National Bureau of Economic Research. Retrieved From-Https://Core.Ac.Uk/Download/Pdf/6402926.Pdf

[35] Heintz, J.(2006). Globalization, economic policy and employment: Poverty and gender implications, ILO Employment Strategy Papers 2006/3 (Geneva, ILO).

[36] Himanshu(2011), “Employment trends in India: A re-examination.” Economic and Political Weekly 46 (37):43–59.

[37] International Labour Organization (ILO); Organisation for the Economic Co-operation and Development(OECD); International Monetary Fund (IMF); and World Bank. 2014. Achieving stronger growth by promoting a more gender-balanced economy, Report prepared for the G20 Labour and Employment

Ministerial Meeting, Melbourne, 10–11 Sep. (Geneva, Paris and Washington, DC)
[38] Kannan, K.P., and G. Raveendran(2012) "Counting and profiling the missing labor force." Economic and Political Weekly 47 (6):77-80.
[39] Kapsos, S., Silbermann, A., & Bourmpoula, E. (2014). Why is female labour force participation declining so sharply in India?. ILO., Retrieved From-http://www.oit.org/wcmsp5/groups/public/---dgreports/--inst/documents/publication/wcms_250977.pdf
[40] Kilby, P. (2011.) .NGOs in India: The Challenges of Women's Empowerment and Accountability.London: Routledge,
[41] Kapil Deep Singh and Jayanty K Sinha (2006), The Indian Economic Association 89th Annual Conference Volume Part – 2, pp. 1070 – 1071
[42] Klasen, S. and D. Schüler ,(2009) "Reforming the Gender-Related Development Index (GDI) and the Gender Empowerment Measure (GEM): Some Specific Proposals", Retrieved from http://www2.vwl.wiso.uni-goettingen.de/ibero/papers/DB186.pdf
[43] 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.
[44] Klugman, J., Hanmer, L., Twigg, S., Hasan, T., McCleary-Sills, J., & Santamaria, J. (2014). Voice and agency: Empowering women and girls for shared prosperity. World Bank Publications. Retrieved From-https://openknowledge.worldbank.org/handle/10986/19036.
[45] Lahoti, R., & Swaminathan, H. (2013). Economic growth and female labour force participation in India.
[46] Luci, A. (2009). Female labour market participation and economic growth. International Journal of Innovation and Sustainable Development, 4(2-3), 97-108. Retrieved From-https://www.researchgate.net/profile/Angela_Greulich_Luci/publication/247835131_Female_labour_market_participation_and_economic_growth/links/55ef299c08ae199d47c00053/Female-labour-market-participation-and-economic-growth.pdf
[47] Mamta Mokta .(2014), Empowerment of Women in India: A critical analysis. Indian Journal Of Public Administration . Vol. Lx, No. 3, July-September 2014.Pg 473-488.
[48] Margaret, S., and Kala,(2013.) N. Study on Impact of NGO Interventions on the Empowerment of Women, Journal of Business Management & Social Sciences Research. 2:3, 1-6, Retrieved from: http://borjournals.com/Research_papers/Mar_2013/1171M.pdf.
[49] Middleton, John & Adrian Ziderman& Arvil Van Adams,(1993) Skill for Productivity: Vocational Education and Training in Developing Countries. New York: Oxford University Press,
[50] Namita Bhandare, (2017), Female Labour Participation In India Is Plummeting In Spite Of Rising Girls' Education: Why And How Do We Fix It? Retrieved From Https://Thewire.In/Labour/Indian-Workplaces-Losing-Women-Nationwide
[51] Namit K Srivastava. Women Empowerment In India. Retrieved From: file:///C:/Users/himanshu/Desktop/Women%20Empowerment%20in%20India%20-%20Rights,%20Laws,%20Policies.html
[52] Nath, K. (1970). Female work participation and economic development: A regional analysis. Economic and Political Weekly, 846-849.
[53] Neff, D. F., Sen, K., & Kling, V. (2012). The Puzzling Decline in Rural Women's Labor Force Participation in India: A Re examination.Organization for Economic Cooperation and Development (OECD) 2012. , Gender Equality in Education, Employment and Entrepreneurship: Final Report to the MCM
[54] Piritta Sorsa (2014) India – igniting inclusive growth by raising female economic participation Retrieved From-https://voxeu.org/article/india-s-female-economic-participation

IJASRM
International Journal of Advanced Scientific Research and Management, Volume 3 Issue 12, Dec 2018
www.ijasrm.com
ISSN 2455-6378

[55] Psacharopoulos, G., and Tzannatos , Z. (1989). Female labor force participation: An international perspective. The International Ban k for Reconstruction and Development /TH E WORL D BANK. Retrieved From http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.828.4549&rep=rep1&type=pdf
[56] Rani, U., & Unni, J. (2009). Do Economic Reforms InfluenceHome-Based Work? Evidence from India. Feminist Economics, 15(3), 191-225.
[57] Sanghi, S., Srija, A., & Vijay, S. S. (2015). Decline in rural female labour force participation in India: A relook into the causes. Vikalpa, 40(3), 255-268. Retrieved From-http://niti.gov.in/writereaddata/files/document_publication/Decline%20in%20Rural%20Female%20Labour%20Force%20Participation%20in%20India.pdf
[58] Sarkar, S., Sahoo, S., & Klasen, S. (2017). Employment Transitions of Women in India: A Panel Analysis., Retrieved From-http://ftp.iza.org/dp11086.pdf
[59] Sen, A.(1987) Gender and cooperative conflicts, WIDER Working Paper No. 18 (Helsinki,United Nations University, World Institute for Development Economics Research (UNU-WIDER)).
[60] Tansel, Aysit.(2002) "Economic Development and Female Labor Force Participation in Turkey: Time-series Evidence and Cross-section Estimates." METU/ERC Working Paper. UNESCAP.(2007) METU/ERC Working Paper. , Retrieved From-http://papers.ssrn.com/sol3/papers.cfm?abstract_id=301946.
[61] The World Bank Policy Rt -21776 , January 2001, Retrieved From-http://documents.worldbank.org/curated/en/512911468327401785/pdf/multi-page.pdf
[62] Tsani, S., Paroussos, L., Fragiadakis, C., Charalambidis, I., & Capros, P. (2012). Female Labour Force Participation and Economic Development in Southern Mediterranean Countries: What Scenarios for 2030?. Retrieved From-http://aei.pitt.edu/59153/1/No_19_Tsani_et_al_Female_Labour_Force_Participation.pdf
[63] UN report (October 16, 2014), Synergies between sustainable development and gender equality undeniable, policy change needed: Focus on women's rights, capabilities, leadership and unpaid care work Date, Retrieved From -http://www.unwomen.org/en/news/stories/2014/10/world-survey-2014-press-release.
[64] World Bank (2001). Engendering Development: Through Gender Equality in Rights, Resources, and Voice. New York and Oxford: Oxford University Press; Washington, D.C.: World Bank.
[65] World Bank (2011). World Development Report 2012: Gender Equality and Development. Washington, D.C.: World Bank.
[66] http://www.oecd.org/employment/50423364.pdf. p. 17. Accessed on 12 October 2018
[67] https://www.thehindubusinessline.com/economy/closing-gender-gap-in-jobs-by-2025-can-add-1-trillion-to-indias-economy-ilo/article9727914.ece-Accessed on 16 October 2018
[68] https://statcompiler.com/en/- Accessed on 18 September 2018
[69] http://www.nber.org/papers/w6321.pdf-Accessed on 21 August 2018

ISSN (Print): 2320-5504
ISSN (Online): 2347-4793

Asia Pacific Journal of Research
A peer reviewed international Journal
IMPACT FACTOR : 6.58
Print -ISSN -2320 - 5504
Online-E-ISSN - 2347 - 4793
University Grants Commission
Approved Journal S.N.45797

EXPERIENCE OF WOMEN IN ENGINEERING EDUCATION: LITERATURE REVIEW

Addissie Melak

Ph.D in Economics research scholar,

Department of Humanity, Delhi Technological University, Delhi -110042

Dr. Seema Singh

Associate professor in Economics,

Department of Humanity, Delhi Technological University, Delhi -110042

ABSTRACT

Despite significant improvements in the last couple of years, women are still under-represented in science and technology, both in the academic and private sector. This is due to a variety of reasons, mostly related to the role allocated to women in modern society as well as pre-existing biases and women's perceptions of commitment to engineering careers. The purpose of this paper is to evaluate experience of women in engineering education in various countries by reviewing research works which is done by researchers related to experience of women in engineering education at different outlooks that can help to have ideas about the trends of women in the field as well as what are the barriers that makes women being underrepresented in engineering education. As such the author concludes that these reviewed papers indicates that there is underrepresentation of women in engineering education in various parts of the world, even if there is some improvement.

Keywords: Women, Engineering Education, Literature review

Introduction

The number of women in science and engineering is growing, yet men continue to outnumber women, especially at the upper levels of these professions. In elementary, middle, and high school, girls and boys take math and science courses in roughly equal numbers, and about as many girls as boys leave high school prepared to pursue science and engineering majors in college. Yet fewer women than men pursue these majors. Among first-year college students, women are much less likely than men to say that they intend to major in science, technology, engineering, or math (STEM). By graduation, men outnumber women in nearly every science and engineering field, and in some, such as physics, engineering, and computer science, the difference is dramatic, with women earning only 20 percent of bachelor's degrees. Women's representation in science and engineering declines further at the graduate level and yet again in the transition to the workplace (Catherine Hill et al., 2010).

Despite significant improvements in the last couple of years, women are still under-represented in science and technology, both in the academic and private sector. This is due to a variety of reasons, mostly related to the role allocated to women in modern society as well as pre-existing prejudices that form glass ceilings while encouraging male presence in the workplace. It is also however, a result of information or lack of, which places young women in difficult position of making a career choice, with little knowledge of

ISSN (Print): 2320-5504
ISSN (Online): 2347-4793

available possibilities. Parents, teachers, and career guidance counselors all have a significant role in assisting or hindering the way young women chose their career paths and that choice begins early on from school, all the way through to higher education (Dimitriadi, 2013).

There has been considerable effort in the last decade to increase the participation of women in engineering through various policies. However, there has been little empirical research on gender disparities in engineering which help underpin the effective preparation, co-ordination, and implementation of the science and technology (S&T) policies (Ghiasi. G et.al. , 2015). Among the almost 2000 students who enrolled in challenge-based engineering course in Texas, females constituted a clear minority, comprising only a total of 14% of students. Quantitative analyses of surveys administered at the beginning of the school year 2011 revealed statistically significant gender gaps in personal attitudes towards engineering and perceptions of engineering climate. Specifically, researcher found that compared to males, females reported lower interest in and intrinsic value for engineering, and expressed less confidence in their engineering skills. Additionally, female students felt that the classroom was less inclusive and viewed engineering occupations as less progressive. Gender disparities on all of these measures did not significantly decrease by the end of the school year (C. Riegle-Crumb, C. Moore, 2013**)**. In the current state of STEM higher education, policies aimed at increasing women's involvement focus only on access. Feminist engineering education is a rare component of any engineering education program at universities across the nation. Less than 20% of undergraduates earning a degree in engineering are women, (Elsa Q. Villa et al., 2016).

The pattern of gender participation in technology is similar in Sweden to what has been reported globally. In the early years during which the modern engineering profession and education was developed and institutionalized, women were excluded from participation. However, in Sweden today, it is seen as good for women to become engineers, even necessary. The proportion of women students in engineering has grown from well below 10% in the early 1970s, to about 30% in the late 1990s to early 2000s. Despite the gains in gender equality, however, it is apparent that ambivalence still surrounds `the woman engineer' today. The proportion of women is high in some programmes and very low in others. The career paths of women and men who are engineers are not comparable either. Moreover, a cultural contradiction between being a woman and being an engineer is mirrored in every-day language, where women are depicted as `soft' while technology is described as `hard' (Maria Udeân, 2002).

The purpose of this paper is to evaluate experience of women in engineering education in various countries by reviewing research works which is done by researchers related to experience of women in engineering education in different counties at different time in different direction.

Literature Review

P.D. Schreuders et al. (2009) investigate Pipeline or personal preference: women in engineering. Although the number of women in the engineering field has increased since the 1960s, those increases have largely stagnated over the last few years. . This study examines based on data collected from engineering students attending American universities using an online survey from the fall of 2005 through the spring of 2007. A total of 969 mane and women students from 21 US universities participated in this online survey. This survey used a convenience sample. Data from the survey were exported directly from Survey Suite into an Excel. In Excel, all string data were converted into numerical data for analysis, then enter into SPSS version 14. The finding indicates that, as educators, no need to worry about the willingness of students to learn engineering skills and that the creation of a gender-balanced curriculum for engineering disciplines can be achieved without sacrificing basic engineering skills. Although no significant gender differences emerged in terms of fundamental engineering activities. Their analysis indicated key gender differences in interest in specific engineering activities.

A.E. Smith and B. Dengiz (2010) studied Women in engineering in Turkey – a large scale quantitative and qualitative examination. Data were collected from Middle East Technical University (METU) (Ankara), Bilkent University (Ankara), Gazi University (Ankara), Hacettepe University (Ankara), Bogazici University (Istanbul), Marmara University (Istanbul), and Istanbul Technical University (Istanbul).A total of 671 female students who were attending these universities responded to the survey. The methods include survey and facilitated focus groups. The study shows that women in Turkey choose engineering mainly because they enjoy the underlying mathematics and science. There is no gender bias on the part of teachers or fellow students; however, women students believe that they have fewer opportunities than male peers and acutely feel the lack of role models. Working professionals in industry or government perceive that women assume a more indirect, supporting role; however, women overall strongly affirm their selection of engineering despite some negative factors.

A.Powell et al. (2011) examine a poisoned chalice? Why UK women engineering and technology students may receive more 'help' than their male peers. The UK engineering and technology (E&T) sector is male-dominated, with women facing various cultural and structural barriers in entering and developing their careers within it. This paper explores women's experiences of higher education in engineering and technology, focusing on the gendered help and support women were found to receive, as well as possible causes and consequences of this behaviour. The researcher adopts a qualitative design, using semi-structured interviews with 53 women engineering and technology students' undergraduate degree at a pre- and post-1992 university. Purposive sampling was undertaken of second-year women E&T undergraduates, since students must decide in this year whether or not to undertake the industrial placement. The gendered help women received is shown to be something of a 'poisoned chalice'; although there are sometimes short-term benefits, this is likely to hinder women from progressing in their careers at the same rate as their male peers.

S. Barnard et al. (2012) investigate they're not girly girls': an exploration of quantitative and qualitative data on engineering and gender in higher education. The quantitative data were collected from each country in the sample – UK, Spain, France, Austria,

ISSN (Print): 2320-5504
ISSN (Online): 2347-4793

Lithuania and Serbia from the year 1972-2008. Through in-depth, semi-structured interviews, in order to uncover engineering students' thoughts and feelings about their educational decisions and experiences in their own words. Interviewers were only from UK in total, 24 interviews with students took place at four universities with 12 female and 12 male students, all within civil engineering departments in 2008. Only students who were in the second, third or final year of their programme were interviewed, as this meant they had more experience to draw on and the questioning would therefore be more fruitful. The interview questions were designed to cover the widest possible range of students' experiences from their decisions to study engineering, their experiences on the programme, reflections on the curriculum taught and gender balance in the staff. The data were analyzed in terms of percentage. The data presented show that, despite some progress in the numbers of women entering engineering programmes, this may be the result of general trends towards increasing numbers of women entering higher education as a whole. European statistics on gender demonstrate a variance between sub-disciplines within engineering and between nations.

Kacey D. Beddoes (2012) examine Feminist Scholarship in Engineering Education: Challenges and Tensions. This article investigate the current state of feminism in the emerging field of engineering education and identifies barriers, challenges, and tensions experienced by scholars and educators who have been involved with feminist engineering education initiatives. Using data from fifteen educator and researchers in-depth interviews conducted through email, in person and Skype selected from US and Australia to show international perspective. The participants worked in several different types of institutions, Snowball sampling was used to select 2 Australian participants and conducted the semi-structured interviews between April 2009 and December 2010. Researcher took a grounded theory approach to analyze the data and identified seven categories of barriers, challenges, and tensions. The categories were: engineering knowledge, culture, and training; engineering education as an emerging research field; publishing; institutional constraints; productive, non-alienating critique and the feminist label; legitimacy and risk; and sustainability.

Singh .S (2013) studied women in engineering education in India**.** The main purpose of the paper was to analyze data on women enrolment as well women faculty members throughout the country. Secondary data from year 1980 – 2005 were used. The researcher concludes that though enrolment in term of number as well as percentage, women participation has increased throughout the country, the growth is not even throughout the country. It varies between states and regions. Though affirmative action as reservation and hostel facility for women students has already been taken, but there is need to expand to enhance scope for affirmative action as fellowship for meritorious students and incentive for women students to join Science, Mathematics and Physics stream at plus two level.

Singh. S, S. Fenton (2014) investigate Women Engineers: A Comparative Study between India and Australia. This study was carryout based on 24 case studies of which fifty –fifty percent case were taken from both countries. Inanition to case studies, researcher used secondary data from the year 1950 – 2011 to investigate trends regarding to participation of women in engineering in the study country. The researcher found that though India and Australia are quite different from each other on so many economic indicators but women participation is low at both places. However, it is important to note that till 1991 it was more or less same but Indian percentage participation has overtaken and also indicates child bearing is the major Couse of attrition. Unless steps are taken to prevent the attrition, there will be a continued loss economically to the country. Researcher inferred that greater responsibility of care work, lake of support system for domestic work within and outside family, male chauvinism among some of the male bosses and opportunity of working part time are some of the reasons for low participation in Australia. While in India, availability of support system, expansion of engineering education system and financial independence at early age are some of the factors which are attracting young girls towards engineering education.

C. García Villa and E.M. González y González (2014) studied Women students in engineering in Mexico: exploring responses to gender differences. The focus of this study was how gender can shape the experiences of female college students in engineering programs. The methodology used in this study was exploratory and descriptive. Participants in this study were female engineering students enrolled in selected colleges of engineering in Mexico at least in third year of collage stay. Twenty respondents from two private and two public universities and from different engineering programs were interviewed. The sample was purposive in design. Ten participants were attending private universities, while the other 10 were students in public universities. Primary data was collected through interview in 2007 and analyzed using the constant comparative method. The percentage of women students in engineering in Mexico is still low compared to the percentage of women enrolled in higher education institutions in the country, which has achieved parity with male enrollment. Findings for this study are presented in two sections. Researcher found the challenges faced by female students in engineering colleges in Mexico, namely, a demanding academic curriculum, and competitive and individualistic environment.

N. Kodate et al. (2014) investigate Paving the way and passing the torch: mentors' motivation and experience of supporting women in optical engineering. This study aims to shed light on both these aspects via the lens of mentors, who are at the coalface of guiding female engineers through their education and subsequent careers. Based on data collected from 25 mentors (8 men and 17 women from 8 countries; USA, Japan, Spain, Mexico, Australia, Canada, Italy and Sweden through interview in August 2011 and October 2013 at the conference held at USA. Data were analyzed using QSR NVivo 9.0 data analysis software. The paper explores their experiences of being mentors, as well as their views on recommended actions for nurturing female engineers. The findings reveal that the primary motivation for becoming a mentor was personal for men and women. Many mentors from countries with relatively lower female labor participation rates perceive their roles as guarantors of their mentees' successful future career paths, and a similar trend can be found in mentors in academia.

ISSN (Print): 2320-5504
ISSN (Online): 2347-4793

Masako Hosaka (2014) Women's experiences in the engineering laboratory in Japan, This qualitative study aims to examine Japanese women undergraduate engineering students' experiences of interacting with departmental peers of the same year in the laboratory setting by using interview data of 32 final-year students at two modestly selective national universities (Western University and Central University) in Japan. Data were collected through questioner and analyzing by using constant comparative methods. Expectation state theory that explains unequal relationship between men and women is used as a framework. Findings suggest that women generally had a discouraging experience while working with their male peers. Specifically, women participated less and lost confidence by comparing with the men who appeared to be confident and competent.

Makarova et al. (2016) investigate why is the pipeline leaking? Experiences of young women in STEM vocational education and training and their adjustment strategies. The paper investigates the perceptions of young women during their vocational education and training (VET) in traditionally male-dominated science, technology, engineering, and mathematic fields by analyzing mechanisms. Both qualitative and quantitative data were collected from semi-structured interviews from71 young women who had chosen a STEM career and who were enrolled in Vocational education and training in Swiss secondary schools in the year 2012 in German. The interview were conducted through telephone. The interviews were transcribed verbatim, checked for errors and analyzed by using the qualitative content analysis. Thus, data were summarized, segmented and patterned by a category scheme that was deductively determined, but also inductively enlarged according to the empirical material (using MAXQDA 11 software). The main concepts of this approach are categories that are defined elements of analysis. The finding indicate different processes of gendering in Vocational education and training, uncovering various mechanisms and symbolic actions which contribute to the reproduction of a masculine culture in STEM professions. Furthermore, the study provides insight into the different strategies young women apply in order to adjust to gender-atypical educational and professional life domains.

Ong, Smith and KO (2017) Counterspaces for Women of Color in STEM Higher Education: Marginal and Central Spaces for Persistence and Success. This article explores the struggles of women of color that threaten their persistence in STEM education and how those struggles lead them to search out or create counter spaces. Data were collected from purposive sample of 39 women and interviewed about their STEM higher education experience. Using a framework of Critical Race Theory (CRT) and intersectionality theory, researcher found that counterspaces can be physical settings, as well as conceptual and ideological.

J. Bossart, N. Bharti (2017) investigate Women in Engineering: Insight into Why Some Engineering Departments Have More Success in Recruiting and Graduating Women. The purpose of this research was to gain a better understanding of why women preferred certain engineering disciplines over others. From the year 2000-2016 of undergraduate engineering department data from the University of Florida (UF) and national averages from the National Science Foundation (NSF) were reviewed to evaluate graduation rates for women in engineering. Linear regression of the data was used to identify trends. In the last 17 years, there has been little change in the overall percentage of women engineering undergraduates, but there is a great disparity between the engineering disciplines. Women earn larger proportions of undergraduate degrees in engineering disciplines where they perceive a societal benefit.

Conclusion

All reviewed papers about experience of women in engineering education is summarized as follows. Although no significant gender differences emerged in terms of fundamental engineering activities, there is key gender differences in interest in specific engineering activities (P.D. Schreuders et al., 2009). Women in Turkey choose engineering mainly because they enjoy the underlying mathematics and science. There is no gender bias on the part of teachers or fellow students; however, women students believe that they have fewer opportunities than male peers and acutely feel the lack of role models (A.E. Smith and B. Dengiz (2010). The gendered help women received is shown to be something of a 'poisoned chalice'; although there are sometimes short-term benefits, this is likely to hinder women from progressing in their careers at the same rate as their male peers (A.Powell et al. (2011).

Despite some progress in the numbers of women entering engineering programmes, this may be the result of general trends towards increasing numbers of women entering higher education as a whole. European statistics on gender demonstrate a variance between sub-disciplines within engineering and between nations (**S.** Barnard et al. (2012). Seven barriers, challenges, and tensions are engineering knowledge, culture, and training; engineering education as an emerging research field; publishing; institutional constraints; productive, non-alienating critique and the feminist label; legitimacy and risk; and sustainability that affects women in engineering education (Kacey D. Beddoes, 2012). Though enrolment in term of number as well as percentage, women participation has increased throughout the country, the growth is not even throughout the country. It varies between states and regions (Singh .S, 2013).

Women participation is low at both places. However, it is important to note that till 1991 it was more or less same but Indian percentage participation has overtaken and also indicates child bearing is the major Couse of attrition (Singh. S, S. Fenton, 2014). The percentage of women students in engineering in Mexico is still low compared to the percentage of women enrolled in higher education institutions in the country, which has achieved parity with male enrollment (C. García Villa, G, 2014). The primary motivation for becoming a mentor was personal for men and women. Many mentors from countries with relatively lower female labor participation rates perceive their roles as guarantors of their mentees' successful future career paths, and a similar trend can be found in mentors in academia (N. Kodate et al., 2014).

Women generally had a discouraging experience while working with their male peers. Specifically, women participated less and lost confidence by comparing with the men who appeared to be confident and competent (Masako Hosaka (2014). Different processes of gendering in Vocational education and training, uncovering various mechanisms and symbolic actions which contribute to the reproduction of a masculine culture in STEM professions. Furthermore, the study provides insight into the different strategies young

ISSN (Print): 2320-5504
ISSN (Online): 2347-4793

women apply in order to adjust to gender-atypical educational and professional life domains (Makarova et al., 2016). Researcher found that counterspaces can be physical settings, as well as conceptual and ideological (Ong, Smith and KO, 2017). In the last 17 years, there has been little change in the overall percentage of women engineering undergraduates, but there is a great disparity between the engineering disciplines. Women earn larger proportions of undergraduate degrees in engineering disciplines where they perceive a societal benefit (J. Bossart, N. Bharti, 2017). Generally, this reviewed papers shows that there is underrepresentation of women in engineering education in various parts of the world, even if there is some improvement.

Reference

1. A. Powell et al. (2011) A poisoned chalice? Why UK women engineering and technology students may receive more 'help' than their male peers, Gender and Education, 23:5, 585-599, DOI: 10.1080/09540253.2010.527826
2. Alice E. Smith & Berna Dengiz (2010) Women in engineering in Turkey – a large scale quantitative and qualitative examination, European Journal of Engineering Education, 35:1, 45-57, DOI: 10.1080/03043790903406345
3. Catherine Hill et al. (2010) Women in Science, Technology, and Engineering, and Mathematics, page xiv, Library of Congress Control Number: 2010901076 ISBN: 978-1-879922-40-2 Published by AAUW Washington, DC 20036 Web: www.aauw.org
4. Carmen García Villa & Elsa M. González y González (2014) Women students in engineering in Mexico: exploring responses to gender differences, International Journal of Qualitative Studies in Education, 27:8, 1044-1061, DOI: 10.1080/09518398.2014.924636
5. C. Riegle-Crumb, C. Moore (2013) Examining Gender Inequality In A High School Engineering Course, American Journal of Engineering Education Special Edition Volume 4, Number 1 pp55-66, http://creativecommons.org/licenses/by/3.0/
6. Dimitriadi (2013) Young women in science and technology: the importance of choice, Journal of Innovation and Entrepreneurship vol. 2:No.5 pp1-14:http://www.innovation-entrepreneurship.com/content/2/1/5
7. Elsa Q. Villa et al.(2016), Engineering Education through the Latina Lens, Engineering Education through the Latina Lens, Journal of Education and Learning; Vol. 5, No. 4; ISSN 1927-5250 E-ISSN 1927-5269, doi:10.5539/jel.v5n4p113
8. Ghiasi G, Larivière V, Sugimoto CR (2015) On the Compliance of Women Engineers with a Gendered Scientific System. PLoS ONE vol.10, No.12, pp1-19: e0145931. doi:10.1371/journal.pone.0145931
9. Jean Bossart, Neelam Bharti (2017) Women in Engineering: Insight into Why Some Engineering Departments Have More Success in Recruiting and Graduating Women, American Journal of Engineering Education Volume 8, Number 2 pp127-140 http://creativecommons.org/licenses/by/3.0/
10. Kacey D. Beddoes (2012) Feminist Scholarship in Engineering Education: Challenges and Tensions, Engineering Studies, 4:3, 205-232, DOI:10.1080/19378629.2012.693932
11. Makarova et al. (2016), why is the pipeline leaking? Experiences of young women in STEM vocational education and training and their adjustment strategies, Empirical Research in Vocational Education and Train (2016) vol.8, No.2 pp1-18: DOI 10.1186/s40461-016-0027-y
12. Maria Udeân (2002) the Impact of Women on Engineering: A Study of Female Engineering Students' Thesis Topics, International Journal of Engineering Education. Vol. 18, No. 4, pp. 458, 2002 0949-149X/91 $3.00+0.00 Printed in Great Britain at TEMPUS Publications
13. Masako Hosaka (2014) Women's experiences in the engineering laboratory in Japan, European Journal of Engineering Education, 39:4, 424-431, DOI: 10.1080/03043797.2014.883363
14. Naonori Kodate, Kashiko Kodate & Takako Kodate (2014) Paving the way and passing the torch: mentors' motivation and experience of supporting women in optical engineering, European Journal of Engineering Education, 39:6, 648-665, DOI: 10.1080/03043797.2014.899323
15. Ong, Smith and KO (2017) Counterspaces for Women of Color in STEM Higher Education: Marginal and Central Spaces for Persistence and Success, Journal of research in science teaching vol. 55, no. 2, pp. 206–245, DOI:10.1002/tea.21417
16. P. D. Schreuders, S. E. Mannon & B. Rutherford (2009) Pipeline or personal preference: women in engineering, European Journal of Engineering Education, 34:1, 97-112, DOI:10.1080/03043790902721488
17. S. Barnard, T. Hassan, B. Bagilhole & A. Dainty (2012) 'They're not girly girls': an exploration of quantitative and qualitative data on engineering and gender in higher education, European Journal of Engineering Education, vol.37:No.2, pp193-204, DOI: 10.1080/03043797.2012.661702
18. Singh. S, S. Fenton (2014) Women Engineers: A Comparative Study between India and Australia International Journal of Advancements in Research & Technology, Volume 3, Issue 7, pp108-121, ISSN 2278-7763
19. Singh.S (2013) Women in Engineering Education in India http://www.youblisher.com/p/109258-Women-in-Engineering-Education-in-India/

OPSEARCH
https://doi.org/10.1007/s12597-018-00350-3

APPLICATION ARTICLE

Improving the consistency of incomplete hesitant multiplicative preference relation

Mamata Sahu[1] · Anjana Gupta[1]

Accepted: 21 December 2018
© Operational Research Society of India 2019

Abstract

As it know, most decision processes are depends on preference relations as in the forms by comparing different alternatives over criteria. In the process of decision making, a decision maker may give his/her judgement by means of hesitant multiplicative preference relation (HMPR) using the scale of $1/9 - 9$ for hesitancy and uncertainty. Shortage of of time, experience and lack of the experts' professional knowledge lead to form an incomplete hesitant multiplicative preference relation. In this paper we have defined a multiplicative transitive property of HMPR that preserves the hesitancy property. Another aim of this paper is to characterise two different approaches to calculate the missing element of incomplete HMPR from the incomplete one. The first approach involves two step, estimating step calculate the initial values of the missing element and the adjusting step minimizing the error. A linear programming model is developed in the second approach to calculate the missing values of incomplete hesitant multiplicative preference relation. The satisfaction degree and the acceptable consistency of complete HMPR is also checked. The whole procedure is explained with a suitable example.

Keywords Multiplicative preference relation · Hesitant multiplicative preference relation · Hesitant multiplicative element · Linear programming problem · Satisfaction degree · Acceptable consistency

✉ Anjana Gupta
anjanagupta@dce.ac.in

Mamata Sahu
mamata_phd2k13@dtu.ac.in

[1] Department of Applied Mathematics, Delhi Technological University, New Delhi, Delhi 110042, India

Published online: 08 January 2019

1 Introduction

At the point when the people settle on a decision, they are regularly hesitant and indecisive for some reason, which makes it hard to achieve a final decision. The trouble of setting up a common membership degree is not on the grounds that have a margin of error in an intuitionistic fuzzy set [1] or interval-valued fuzzy set [19] or in some possibility distribution(as in type-2 fuzzy sets) with respect to the possibility values, however it is on account of we have an arrangement of possible values [8]. For a better description of this circumstance, Torra [8] proposed the hesitant fuzzy set, which allow the membership degree of an element represented by several possible values lies between 0 and 1. Not only HFS is an extension of fuzzy set that address uncertainty, but also it is an important tool which represents the case in which several membership functions for a fuzzy sets are possible.

In the previous couple of decades, various research have been done on the utilization of preference relations and accordingly different types of preference relation have been developed such as fuzzy preference relation (FPR) [4, 7], multiplicative preference relation (MPR)[6], intuitionistic fuzzy preference relation (IFPR) [16, 17], intuitionistic multiplicative preference relation (IMPR) [10], the interval-valued intuitionistic fuzzy preference relation (IVI-FPR) [14, 21], interval-valued intuitionistic multiplicative preference relation (IVI-MPR) [22], linguistic preference relation [2, 3] etc. In case of fuzzy scenario the scale of membership is taken 0–1 ratio scale and in case of MPR the ratio scale is $1/9-9$.

In real life situation, the preference relation as mentioned above does not explore all promising estimation of preferences of the decision maker because all these preference relations do not think about hesitant fuzzy data while making a pairwise comparison. However, sometimes, to get a more reasonable decision result, a decision organization, which contains a lot of decision-makers (or experts), is authorized to provide the preferences by comparing each pair of alternatives using 0–1 ratio scale, and while providing the degrees to which an alternative is superior to another. It may be the case that, the decision maker is not very much sure about an exact value but has hesitancy between several possible values. In such cases, these several possible values can be considered as an HFE, and a hesitant fuzzy preference relation is constructed when all the preferences over a set of alternatives are provided. It is to note that the development of hesitant fuzzy preference relation is based on the fuzzy preference relation, whose values are expressed by using the 0–1 ratio scale which is uniformly and symmetrically distributed around 0.5. Zhu and Xu [28] define hesitant fuzzy preference relations(HFPRs), and they gave some properties on HFPR.

In numerous genuine issues, nonetheless, the grade of the preferences are not consistently and symmetrically distributed around some value. For example, While getting the ventures of similar assets, an organization with bad performance improves more rapidly than an organization with good performance; to be specific, the distance grades expressing good information should be greater than the grades expressing bad performance. To address such a circumstance,

Saaty's 1–9 scale [6] is a helpful tool, particularly in multiplicative preference relation that has been applied in different areas [11]. To express the preferences with respect to two options Xia and Xu [11] utilized the asymetrical 1–9 scale as opposed to the symmetrical 0.1–0.9 scale in the HFPR and after that they presented another idea of the hesitant multiplicative preference relation (HMPR). Referencing the above example, the experts in the decision making problem do not like to utilize the qualities somewhere in the range of 0 and 1 yet would want to utilize Saaty's ratio scale to provide the degree to which the alternatives x_i is superior to x_j. for example, some experts in the decision-making problem provide 1/3, some provide 4, and some provide 5, then the decision information to which the alternative x_i is superior to x_j can be represented by a set of possible value as $\{1/3, 4, 5\}$ which is called hesitant multiplicative element (HME) [11]. Xia and Xu [11] used the 1–9 scale and introduce the concept of hesitant multiplicative preference relation (HMPR). In this work, we concentrate only on hesitant multiplicative preference relation (HMPR).

1.1 Motivation

In the above mentioned preference relations, it is assumed that the experts are always able to provide information on each entry in a preference relation matrix. It may not be the case forever. Here and there, decision-maker might not have a decent comprehension of a specific query, thus he/she may not make a direct comparison between each two objects; in this manner it is once in a while important to enable the decision-maker to avoid some questionable comparisons flexibly. In this case, incomplete preference relations are obtained, and the whole process may slow down. The above mentioned preference relations are work on complete preference relation. At each level of the decision-making process, decision maker need $\frac{n\times(n-1)}{2}$ judgements to present a complete preference relation of order $n \times n$ in its entire upper or lower triangular portion. Sometimes, it is very difficult to obtain a complete preference relation, for the shortage of time pressure, and lack of sufficient knowledge, in many practical situations, such as medical diagnosis, personal examination. The situation results in decision maker constructs an incomplete preference relation in which some elements are missing.

Consistency plays an important role in decision-making process. As consistency of preference relations influences the final ranking order of alternatives. A much amount of research on consistency on HFPR has been done in the literature. Zhang et al. [24], defined three different additive consistency index to measure the consistency level of an HFPR such as best additive consistency index, the worst additive consistency index, the average additive consistency index. They developed some mixed 0–1 linear programming problem to improve the additive consistency of an HFPR and extended the work for incomplete HFPR. Zhang et al. [25] used the logarithmic least square method and proposed a method to derive priority weight vector from an incomplete HFPR. Also they define the consistency index of an incomplete HFPR to calculate the average deviation between the priority weight vector and the elements of incomplete HFPR. Zhang et al. [25] developed an algorithm to improve

the consistency. Xu et al. [12] define the concept of multiplicative consistency and additive consistency of incomplete HFPR. Using the both additive and multiplicative consistency, they have proposed two goal programming models to derive the priority weights from an incomplete HFPR.

The role of significance of consistency measure for HMPR make sure that the decision-makers are neither arbitrary nor unreasonable.Our work focuses on hesitant multiplicative preference relation. In this paper we have defined the multiplicative transitive property of HMPRs that preserve the hesitancy property of HMPRs. By using this property we have developed an algorithm to construct a complete HMPRs from incomplete one and developed an optimization model to minimize the error. Also we have developed a linear programming problem to calculate the missing value of HMPR. Zhu and Xu [27] developed a method to measure the satisfaction degree of the HMPRs. We have checked the satisfaction degree of new complete HMPRs. Algorithm is illustrated via suitable example with the satisfaction degree more than 80%. Also we have checked the acceptable consistency of HMPR.

The rest of the paper is organized as follows. In Sect. 2, some basic concepts of MPRs and HMPRs are defined briefly. Multiplicative transitive property for HMPRs defined that is used to construct a complete HMPRs from the incomplete one discussed in Sect. 3. A linear programming model is developed in Sect. 4 to find out the missing element of incomplete HMPR directly. Satisfaction degree and acceptable consistency of complete HMPR is also checked in this section . An example is given to illustrate the whole procedure. Some concluding remarks are given in the last section.

2 Preliminaries

In this section we discuss some basic concept of multiplicative preference relation (MPR), hesitant multiplicative set and hesitant multiplicative preference relations.

2.1 Multiplicative preference relations

Let $X = \{x_1, x_2, ..., x_n\}$ be a discrete set of alternatives in a decision making problem which a decision maker has to compare in pairs, and $N = \{1, 2, ..., n\}$ be the set of indices.

Definition 1 [5, 6] An Multiplicative preference relation (MPR) over X is defined by $P = [p_{ij}]_{n\times n}$ where p_{ij} is the multiplicative preference value to which an alternative x_i is preferred to an alternative x_j satisfying

$$p_{ij}p_{ji} = 1, p_{ij} \geq 0, \forall\ i,j \in N.$$

$p_{ij} \in [1/9, 9]$, $\forall i,j = 1, 2, \ldots n$. i.e., multiplicative preference value measured using 1–9 ratio scale. The preference value p_{ij} with value 1 specify indifference between x_i and x_j, $p_{ij} = 9$ specify that the alternative x_i is absolutely preferred to x_j, $p_{ij} = 1/9$ specify that x_j is absolutely preferred to x_i. $p_{ij} > 1$ indicates that the preference value

of the alternatives x_i over x_j and $p_{ij} < 1$ specify the preference value of the alternatives x_j over x_i.

Definition 2 [5, 6] An MPR $P = (p_{ij})_{n\times n}$ satisfying the multiplicative transitivity property $p_{ij} = p_{ik}p_{kj} \forall i,j,k = 1, 2, \ldots, n$ is called multiplicative consistent.

If $P = (p_{ij})_{n\times n}$ is consistent MPR, then it should satisfy the property as

$$p_{ij} = \frac{w_i}{w_j},\ i,j \in N \tag{1}$$

where $w = (w_1, w_2, \ldots, w_n)$ is the priority vector of the objectives satisfying the $\sum_{i=1}^{n} w_i = 1, w_i \geq 0,\ i \in N$

2.2 Hesitant multiplicative set

Xia and Xu [11] defined the concept of hesitant multiplicative set in which elements are the set of possible values.

Definition 3 [11] Hesitant multiplicative set on X is defined by $M = \{(x, b_M(x))|x \in X\}$, where $b_M(x)$ is a set of some values that lies between 1/9 to 9. For convenience $b_M(x) = b_e$ is called the hesitant multiplicative element (HME).

Given an HME $b_e = \{b_e^{\sigma(s)}|s = 1, 2, \ldots, l_{b_e}\}$ where l_{b_e} denotes the number of elements in the HME b_e. $\sigma(s)$ denotes the position of HME.

Let $b_e^{\sigma(s)}$, $b_{e_1}^{\sigma(s)}$, $b_{e_2}^{\sigma(s)}$, are the s^{th} position value of HME b_e, b_{e_1}, and b_{e_2} respectively. Additionally suppose that $l_{b_{e_1}} = l_{b_{e_2}} = l$, or else we can extend the smaller one by adding the numerical values given as the following definition.

Definition 4 [26] Let b_e^+ and b_e^- are the maximum and minimum elements in HME $b_e = \{b_e^{\sigma(s)}|s = 1, 2, \ldots, l_{b_e}\}$ respectively and δ be an optimized parameter that lies between 0 and 1 which can be chosen by the decision makers according to their own risk preferences, then $(b_e^+)^{\delta} \times (b_e^-)^{1-\delta}$ is called an added element that is denoted by $\bar{b}_e$.

Particularly the added element b_e^+ and b_e^- respectively derived from the conditions $\delta = 1$ and $\delta = 0$ that corresponds to the optimism and pessimism rules are taken by the decision makers.

2.3 Hesitant multiplication preference relation

Xia and Xu [11] developed the concept of HMPRs on the basis of hesitant multiplicative elements (HMEs) and multiplicative preference relations (MPRs) given as follows

Definition 5 [11] Hesitant multiplicative preference relation B_R over X is defined by $B_R = (b_{e_{ij}})_{n\times n}$ which is subset of $X \times X$, where $b_{e_{ij}} = \{b_{e_{ij}}^{\sigma(s)} | s = 1, 2, \ldots, l_{b_{e_{ij}}}\}$ is a HME indicating all the possible preference value to which an alternative x_i is preferred to alternative x_j satisfying

$$b_{e_{ij}}^{\sigma(s)} \times b_{e_{ji}}^{\sigma(l_{b_{e_{ji}}}-s+1)} = 1,\ b_{e_{ii}} = \{1\},\ l_{b_{e_{ij}}} = l_{b_{e_{ji}}},\ \forall i, j \in N. \tag{2}$$

where $b_{e_{ij}}^{\sigma(s)}$ is the s^{th} position element in $b_{e_{ij}}$. Xia and Xu [11] mentioned that HME are assumed in increasing order. But it may be the case that some decision makers give their preference in decreasing order that also satisfy the hesitancy property Eq. 2. Also Zhang and Guo [23] proved following theorem

Theorem 1 *Let $B = (b_{e_{ij}})_{n\times n}$, where $b_{e_{ij}} = \{b_{e_{ij}}^{\sigma(s)} | s = 1, 2, \ldots, l_{b_{e_{ij}}}\}$ be an HMPR. It can be transfer into HFPR $H = (h_{e_{ij}})_{n\times n}$, $h_{e_{ij}} = \{h_{e_{ij}}^{\sigma(s)} |\ s = 1, 2, \ldots, l_{h_{e_{ij}}}\}$ by using the transformation $h_{e_{ij}}^{\sigma(s)} = \frac{b_{e_{ij}}^{\sigma(s)}}{1+b_{e_{ij}}^{\sigma(s)}}$, then transfer the HFPR H into the HMPR $B' = (b'_{e_{ij}})_{n\times n}$ by using the transformation $b'_{e_{ij}} = \frac{h_{e_{ij}}^{\sigma(s)}}{1-h_{e_{ij}}^{\sigma(s)}}$, where $b'_{e_{ij}} = \{b'^{\sigma(s)}_{e_{ij}} |\ s = 1, 2, \ldots, l_{b'_{e_{ij}}}\}$,* then $B = B'$.

In our case we are taking the HME into increase/decrease that also satisfy the Theorem 1.

In this paper, we are not considering the case that HMEs are arranged in increasing order always.

For example let the HMPR, B_R as shown below

$$B_R = \begin{bmatrix} \{1\} & \left\{\frac{1}{6}, \frac{1}{5}, \frac{1}{4}\right\} & \{3\} \\ \{4,5,6\} & \{1\} & \{4,2\} \\ \left\{\frac{1}{3}\right\} & \left\{\frac{1}{2}, \frac{1}{4}\right\} & \{1\} \end{bmatrix}$$

Since length of the preference degree of alternative x_i over x_j in HMPR matrix are not same, based on Definition 4, Zhang et al. [26] used the optimized parameter δ to add some elements to an HMPR and finally obtained a normalized hesitant multiplicative preference relation (N-HMPR) defined as follows

Definition 6 [26] Let $b_{e_{ij}}^+$ and $b_{e_{ij}}^-$ be the maximum and minimum elements of HMPR, $B_R = (b_{e_{ij}})_{n\times n}$ be respectively, for $i, j = 1, 2, \ldots, n$, and let δ be an optimized parameter. According to Definition 4 if we add some elements $\bar{b}_{e_{ij}} = (b_{e_{ij}}^+)^{\delta} \times (b_{e_{ij}}^-)^{(1-\delta)}$ to $b_{e_{ij}}$ if $i \le j$ and add some elements $\bar{b}_{e_{ji}} = (b_{e_{ji}}^+)^{(1-\delta)} \times (b_{e_{ji}}^-)^{\delta}$ to $b_{e_{ji}}$ if $i \le j$, then the normalized hesitant multiplicative preference relation (N-HMPR) $\bar{B}_R = \{\bar{b}_{e_{ij}}^{\sigma(s)} | s = 1, 2, \ldots, l_{\bar{b}_{e_{ij}}}\}$ satisfying

$$l_{\bar{b}_{e_{ij}}} = \max\left\{l_{b_{e_{ij}}} | i, j = 1, 2, \ldots, n, i \ne j\right\},\ \bar{b}_{e_{ij}}^{\sigma(s)} \times \bar{b}_{e_{ji}}^{\sigma(l_{\bar{b}_{e_{ji}}}-s+1)} = 1,\ \bar{b}_{e_{ii}} = \{1\}. \tag{3}$$

where $\bar{b}_{e_{ij}}^{\sigma(s)}$ and $\bar{b}_{e_{ji}}^{\sigma(s)}$ is the s^{th} position element in $\bar{b}_{e_{ij}}$ and $\bar{b}_{e_{ji}}$ respectively.

The N-HMPR $\bar{B}_R$ (*let* $\delta = 1$) of the HMPR B_R from the above example as shown in below

$$B_R = \begin{bmatrix} \{1\} & \left\{\frac{1}{6}, \frac{1}{5}, \frac{1}{4}\right\} & \{3,3,3\} \\ \{4,5,6\} & \{1\} & \{4,4,2\} \\ \left\{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right\} & \left\{\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right\} & \{1\} \end{bmatrix}$$

Some times it is very difficult to obtain a complete preference relation especially for the preference relation with higher order. It may be the case that an expert is unable to express his/her opinion over other pairs of alternatives, due to limitation of decision makers' proficient knowledge, experience or time pressure, the provided preference value in HMPR becomes incomplete. Then the incomplete HMPR is constructed in which some elements are missing. The incomplete HMPR is given by Zhang and Guo [23].

Definition 7 (Incomplete HMPR) [20, 23] The HMPR $B_R = (b_{e_{ij}})_{n\times n}$ where $b_{e_{ij}} = \{b_{e_{ij}}^{\sigma(s)} | s = 1, 2, \ldots, l_{b_{e_{ij}}}\}$ is called an incomplete HMPR, if some elements in it are missing that is denoted by $\}\}\ *''$ and the other elements are given by decision makers should satisfy the condition of Definition 5.

Following is an example of incomplete HMPR,

$$\begin{bmatrix} \{1\} & \left\{\frac{1}{6}, \frac{1}{5}\right\} & * \\ \{5,6\} & \{1\} & \{2,4\} \\ * & \left\{\frac{1}{4}, \frac{1}{2}\right\} & \{1\} \end{bmatrix}$$

where $\}\}\ *''$ is the missing element. Based on Definition 6, we can use the optimized parameter δ to add some elements to an incomplete HMPR and obtained an incomplete normalized HMPR defined as follows.

Definition 8 (Incomplete N-HMPR) The incomplete HMPR is called incomplete N-HMPR, if the known elements satisfy the Eq. 3.

Zhang et al. [26] mentioned that if $B = (b_{e_{ij}})_{n\times n}$, $b_{e_{ij}} = \{b_{e_{ij}}^{\sigma(s)} | s = 1, 2, \ldots, l_{b_{e_{ij}}}\}$ be a HMPR and $\bar{B} = (\bar{b}_{e_{ij}})_{n\times n}$, $\bar{b}_{e_{ij}} = \{\bar{b}_{e_{ij}}^{\sigma(s)} | s = 1, 2, \ldots, l, \quad (l = max\{l_{b_{e_{ij}}} | i, j = 1, 2, \ldots n, i \ne j\})$ be its N-HMPR with the optimized parameter δ, then l number of MPR $P^{(s)} = (p_{ij}^{(s)})_{n\times n}, (s = 1, 2, \ldots, l)$ are obtained where,

$$p_{ij} = \begin{cases} \bar{b}_{ij}^{\sigma(s)} & i < j \\ 1 & i = j \\ \bar{b}^{\sigma(l-s+1)} & i > j \end{cases} \tag{4}$$

In 2014 Zhang et al. [26] defined the consistent H-MPR.

Definition 9 [26] Let $B = (b_{e_{ij}})_{n\times n}$ be HMPR and $\bar{B} = (\bar{b}_{e_{ij}})_{n\times n}$ be its N-HMPR with optimized operator δ. HMPR is said to be consistent if all the MPRs obtained from $\bar{B}$ using 4 are consistent.

Definition 10 [26] Let $B = (b_{e_{ij}})_{n\times n}$ be HMPR and $\bar{B} = (\bar{b}_{e_{ij}})_{n\times n}$ be its N-HMPR with optimized operator δ. HMPR is acceptable consistent if all the MPRs obtained from $\bar{B}$ are acceptable consistent.

3 Constructing complete HMPR from an incomplete HMPR

In this section, before discussing about constructing complete HMPR from an incomplete HMPR, we have defined the following consistency property.

Definition 11 Let $B_R = (b_{e_{ij}})_{n\times n}$ and $\bar{B}_R = (\bar{b}_{e_{ij}})$ be HMPR and its N-HMPR respectively on X with an optimization operator δ, satisfying the following multiplicative transitive property

$$\bar{b}_{e_{ij}}^{\sigma(s)} = \bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)} \times \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)} \quad \forall\, i,j,k \in N, \text{ and } i \neq k \neq j,\ s = 1,2,\ldots,l_{\bar{b}_{e_{ij}}}. \tag{5}$$

where $\bar{b}_{e_{ij}}^{\sigma(s)}$, $\bar{b}_{e_{ik}}^{\sigma(s)}$ and $\bar{b}_{e_{kj}}^{\sigma(s)}$ are the s^{th} position elements in $\bar{b}_{e_{ij}}$, $\bar{b}_{e_{ik}}$ and $\bar{b}_{e_{kj}}$ respectively then the HMPR, B_R is said to be a multiplicative consistent HMPR with optimization parameter δ. The multiplicative transitive property of HMPRs satisfy the property of hesitancy i.e., $\bar{b}_{e_{ij}}^{\sigma(s)} \times \bar{b}_{e_{ji}}^{\sigma(l_{\bar{b}_{e_{ji}}}-s+1)} = 1$.

Definition 12 Let $B_R = (b_{e_{ij}})_{n\times n}$ and $\bar{B}_R = (\bar{b}_{e_{ij}})$ be HMPR and its N-HMPR on X with an optimization operator δ, Then the HMPR is called a consistent H-MPR if

$$\bar{b}_{e_{ij}}^{\sigma(s)} = \left\{ \prod_{k=1}^{n} \left(\bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)} \times \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)} \right) \right\}^{1/n} \tag{6}$$

The multiplicative consistency property of HMPR will take part as an important role to construct complete HMPRs from incomplete HMPRs. We define a multiplicative consistent incomplete HMPR.

Definition 13 The incomplete HMPR, $B_R = (b_{e_{ij}})_{n\times n}$ is called multiplicative consistent if all the known elements satisfy the conditions of Eq. 5.

Zhang et al. [26] proposed the definition of acceptable incomplete HMPR.

Definition 14 [26] Incomplete HMPR is said to be acceptable incomplete HMPR if at least one known element present in each row or each column of HMPR.

Before going to the next section we have defined an algorithm for constructing a complete HMPR from an incomplete one.

Algorithm 1

1. Consider an incomplete HMPR, B_R.
2. Normalize the incomplete HMPR using Definition 6 is denoted by $\bar{B}_R = \bar{b}_{e_{ij}}$.

$$\bar{b}_{e_{ij}} = \begin{cases} * & \text{if } \bar{b}_{e_{ij}} \notin \Omega; \\ \bar{b}_{e_{ij}} & \text{if } \bar{b}_{e_{ij}\in\Omega}. \end{cases} \tag{7}$$

where Ω is the set of known element and '$*$' is the value of the missing element.

3. Initially the missing element can be calculated using the formula

$$\bar{b}_{e_{ij}}^{\sigma(s)(*(0))} = \left\{ \prod_{k\in T_{ij}} \left(\bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)} \times \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)} \right) \right\}^{1/t_{ij}} \tag{8}$$

where t_{ij} is the number of the elements of the set $T_{ij} = \{k | \bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)}, \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)} \in \Omega\}$ which indicates that there exist different pairs of adjoining known elements to calculate the missing element $\bar{b}_{e_{ij}}$. It is note that, the subsequent missing elements are evaluated by using the previous calculated missing values.

4. The initial values are obtained.

Sometimes initial values obtained may not lie in between 1 / 9 to 9 or does not satisfy the required condition of hesitancy as well as consistency. For solving this type of difficulties, we have developed an optimization model that minimizes the error. MATLAB optimization tool box is used to solve the optimization model.

The multiplicative transitive property (Eq.5) of H-MPR can be rewritten as

$$\log \bar{b}_{e_{ij}}^{\sigma(s)} = \log \bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)} + \log \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)} \quad \forall\, i,j,k \in N, \text{ and } i \neq k \neq j,\ s = 1,2,\ldots,l_{\bar{b}_{e_{ij}}}. \tag{9}$$

In more general inconsistent cases, we can utilize Eq. 9 to minimize the error. To do this we have constructed the following optimization model.

MODEL (M)

$$\text{Min} \quad \sum_{(i,j)\notin\Omega} \varepsilon_{\bar{b}_{e_{ij}^k}}$$

subject to

$$\varepsilon_{\bar{b}_{e_{ij}^k}} - |\log \bar{b}_{e_{ij}}^{\sigma(s)} - \left(\log \bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)} + \log \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)}\right)| = 0,\ i \neq k \neq j,\ s = 1,2 \ldots l_{\bar{b}_{e_{ij}}}$$

$$1/9 - \bar{b}_{e_{ij}}^{\sigma(s)} \leq 0, s = 1,2 \ldots l_{\bar{b}_{e_{ij}}}$$

$$\bar{b}_{e_{ij}}^{\sigma(s)} - 9 \leq 0, s = 1,2 \ldots l_{\bar{b}_{e_{ij}}}$$

$$\bar{b}_{e_{ij}}^{\sigma(s)(0)} = \bar{b}_{e_{ij}}^{\sigma(s)(*(0))}, s = 1,2 \ldots l_{\bar{b}_{e_{ij}}}$$

$$\bar{b}_{e_{ij}}^{\sigma(s)} \times \bar{b}_{e_{ji}}^{\sigma(l_{\bar{b}_{e_{ij}}}-s+1)} = 1$$

where $\bar{b}_{e_{ij}}^{\sigma(s)(*(0))}$ are the intial value obtained from the Algorithm 1.

In the next section we have developed a linear programming model to calculate the missing element of incomplete HMPR.

4 Linear programming model to estimate the missing values of incomplete HMPR

Many researcher [9, 10, 13, 15, 18] noted preference relation may incomplete for various reason. To cope with incomplete hesitant multiplicative preference relation, in this section we have constructed a linear programming model to evaluate the missing element. This model can address the situation where ignorance exist. Let $\bar{B}_R = (\bar{b}_{e_{ij}})_{n\times n}$ ba an incomplete normalized hesitant multiplicative preference relation is called consistent then Eq. 6 is satisfied.

From Definition 5, we know that $\bar{B}_R$ is consistent if and only if Eq. 6 holds for the strictly upper triangular elements, namely

$$\bar{b}_{e_{ij}}^{\sigma(s)} = \left\{\prod_{k=1}^{n}\left(\bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)} \times \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)}\right)\right\}^{\frac{1}{n}} \tag{10}$$
$$\forall\ i,j \in N,\ \text{and}\ i \neq k \neq j,\ i<j,\ s = 1,2,\ldots,l_{\bar{b}_{e_{ij}}}.$$

Generally Eq. 10 does not hold. Because the missing values make the consistency of incomplete hesitant multiplicative preference relations the higher the better, we relax and approximate Eq. 10 by minimizing the error ε_{ij}, where

$$\varepsilon_{ij} = \left|\bar{b}_{e_{ij}}^{\sigma(s)} - \left\{\prod_{k=1}^{n}\left(\bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)} \times \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)}\right)\right\}^{\frac{1}{n}}\right| \tag{11}$$
$$\forall\ i,j \in N,\ \text{and}\ i \neq k \neq j,\ i<j,\ s = 1,2,\ldots,l_{\bar{b}_{e_{ij}}}.$$

ε_{ij} is a constant when all the values of Eq. 11 are known. Therefore, here we only consider the equations having unknown values. To do this, Let Ω is the set of all known element of hesitant multiplicative preference relations and

$$\delta_{ij} = \begin{cases} 1 & k \in \Omega \\ 0 & \text{otherwise} \end{cases}$$

Define

$$\varepsilon_{ij} = \delta_{ij}\left|\bar{b}_{e_{ij}}^{\sigma(s)} - \left\{\prod_{k=1}^{n}\left(\bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)} \times \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)}\right)\right\}^{\frac{1}{n}}\right| \tag{12}$$
$$\forall\ i,j \in N,\ \text{and}\ i \neq k \neq j,\ i<j,\ s = 1,2,\ldots,l_{\bar{b}_{e_{ij}}}.$$

Thus, we can construct the following multi-objective programming model to estimate the missing values:

$$\text{(MOP)}\quad \min \varepsilon_{ij} = \delta_{ij}\left|\bar{b}_{e_{ij}}^{\sigma(s)} - \left\{\prod_{k=1}^{n}\left(\bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)} \times \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)}\right)\right\}^{\frac{1}{n}}\right|$$
$$\forall\ i,j \in N,\ \text{and}\ i \neq k \neq j,\ i<j, \qquad s = 1,2,\ldots,l_{\bar{b}_{e_{ij}}}$$
$$\text{s.t}\ 1/9 \leq \bar{b}_{e_{ij}}^{\sigma(s)} \leq 9,\ \bar{b}_{e_{ij}}^{\sigma(s)} \in \Gamma,$$

where $\Gamma = \{\bar{b}_{e_{ij}}^{\sigma(s)} | \bar{b}_{e_{ij}}^{\sigma(s)} \text{is a missing value}, i,j \in N,\ i<j,\ s = 1,2,\ldots,l_{\bar{b}_{e_{ij}}}\}$

The solution of the above minimization problem can be obtained by solving the following goal programming model

$$\text{(OP)} \sum_{i,j=1,i<j}^{n}\ \sum_{k\in\Omega,s=1,2,\ldots,l_{\bar{b}_{e_{ij}}}} \left(d_{ij,k}^{\sigma(s)+} + d_{ij,k}^{\sigma(s)-}\right)$$

$$s.t \begin{cases} \delta_{ij}\left(\bar{b}_{e_{ij}}^{\sigma(s)} - \left\{\prod_{k=1}^{n}\left(\bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)} \times \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)}\right)\right\}^{\frac{1}{n}}\right) - d_{ij,k}^{\sigma(s)+} + d_{ij,k}^{\sigma(s)-} = 0;\ i,j \in N,\ i<j,\ s = 1,2,\ldots,l_{\bar{b}_{e_{ij}}} \\ 1/9 \leq \bar{b}_{e_{ij}}^{\sigma(s)} \leq 9 \qquad \bar{b}_{e_{ij}}^{\sigma(s)} \in \Gamma \\ d_{ij,k}^{\sigma(s)+}, d_{ij,k}^{\sigma(s)-} \geq 0 \qquad i,j \in N,\ i<j,\ s = 1,2,\ldots,l_{\bar{b}_{e_{ij}}} \end{cases}$$

where $d_{ij,k}^{\sigma(s)+} = \left(\bar{b}_{e_{ij}}^{\sigma(s)} - \left\{\prod_{k=1}^{n}\left(\bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)} \times \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)}\right)\right\}^{\frac{1}{n}}\right) \vee 0$ and

$$d_{ij,k}^{\sigma(s)-} = \left(\left\{ \prod_{k=1}^{n} \left(\bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)} \times \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)} \right) \right\}^{\frac{1}{n}} - \bar{b}_{e_{ij}}^{\sigma(s)} \right) \vee 0 \ \forall \ i,j \ \in N, \ i<j, \ s=1,2,\ldots,l_{\bar{b}_{e_{ij}}}.$$

Using the model (OP) one can obtain the missing values. Since the above goal programming model (OP) is nonlinear, the optimal solution is varied with the initial feasible solution. Thus $z_{ij} = \ln \bar{b}_{e_{ij}}^{\sigma(s)}$ for all $i,j \ \in N, \ i<j, \ s=1,2,\ldots,l_{\bar{b}_{e_{ij}}}$.

Then Eq. 10 can be transformed into

$$\begin{aligned} z_{ij} = \ln \bar{b}_{e_{ij}}^{\sigma(s)} &= \ln \left\{ \prod_{k=1}^{n} \left(\bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)} \times \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)} \right) \right\}^{\frac{1}{n}} \\ &= \frac{1}{n} \ln \left\{ \prod_{k=1}^{n} \left(\bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)} \times \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)} \right) \right\} \\ &= \frac{1}{n} \sum_{k=1}^{n} \ln \left(\bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)} \times \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)} \right) \\ &= \frac{1}{n} \sum_{k=1}^{n} \left(\ln \bar{b}_{e_{ik}}^{\sigma(l_{\bar{b}_{e_{ik}}}-s+1)} + \ln \bar{b}_{e_{kj}}^{\sigma(l_{\bar{b}_{e_{kj}}}-s+1)} \right) \\ &= \frac{1}{n} \sum_{k=1}^{n} (z_{ik} + z_{kj}) \end{aligned} \tag{13}$$

Then we construct a linear programming model

$$\text{(LP)} \quad \sum_{i,j=1,i<j \ k\in\Omega,s=1,2,\ldots,l_{\bar{b}_{e_{ij}}}}^{n} \left(d_{ij,k}^{\sigma(s)+} + d_{ij,k}^{\sigma(s)-} \right)$$

$$s.t \begin{cases} \delta_{ij}\left(z_{ij} - \frac{1}{n}\sum_{k=1}^{n}(z_{ik}+z_{kj}) \right) - d_{ij,k}^{\sigma(s)+} + d_{ij,k}^{\sigma(s)-} = 0, & i,j \ \in N, \ i<j \ s=1,2,\ldots,l_{\bar{b}_{e_{ij}}} \\ 1/9 \le z_{ij} \le 9 & z_{ij} \ \in \Gamma' \\ d_{ij,k}^{\sigma(s)+}, d_{ij,k}^{\sigma(s)-} \ge 0 & i,j \ \in N, \ i<j \ s=1,2,\ldots,l_{\bar{b}_{e_{ij}}} \end{cases}$$

where $\Gamma' = \{z_{ij} | z_{ij} \text{is a missing value}, i,j \ \in N, \ i<j\}$, $d_{ij,k}^{\sigma(s)+} = \left(z_{ij} - \frac{1}{n}\sum_{k=1}^{n} (z_{ik}+z_{kj}) \right) \vee 0$ and

$$d_{ij,k}^{\sigma(s)-} = \left(\frac{1}{n}\sum_{k=1}^{n}(z_{ik}+z_{kj}) - z_{ij} \right) \vee 0 \ \forall i,j \ \in N, \ i<j \ s=1,2,\ldots,l_{\bar{b}_{e_{ij}}}.$$

To illustrate the above two model i.e., optimization model (Model M) and linear programming model (LP), we have taken two examples, which is discussed in below.

Example 1 Consider a decision making problem with five sets of alternatives $x_i = \{x_1, x_2, x_3, x_4, x_5\}$. The decision maker judge these five alternatives by pairwise comparison and provides his/her judgement given in a matrix denoted by B_R given as follows:

$$B_R = \begin{bmatrix} \{1\} & \left\{\frac{7}{4},\frac{5}{6},\frac{1}{3}\right\} & \{3,5,7\} & * & \left\{\frac{3}{5},\frac{4}{5}\right\} \\ \left\{3,\frac{6}{5},\frac{4}{7}\right\} & \{1\} & \{4,6,9\} & * & \left\{\frac{16}{35},\frac{24}{25},\frac{18}{10}\right\} \\ \left\{\frac{1}{7},\frac{1}{5},\frac{1}{3}\right\} & \left\{\frac{1}{9},\frac{1}{6},\frac{1}{4}\right\} & \{1\} & * & * \\ * & * & * & \{1\} & \left\{\frac{1}{4},\frac{1}{4},\frac{1}{4}\right\} \\ \left\{\frac{5}{4},\frac{5}{3}\right\} & \left\{\frac{10}{18},\frac{25}{24},\frac{35}{16}\right\} & * & \{4,4,4\} & \{1\} \end{bmatrix}$$

The NH-MPR $\bar{B}_R$ is given by

$$\bar{B}_R = \begin{bmatrix} \{1\} & \left\{\frac{7}{4},\frac{5}{6},\frac{1}{3}\right\} & \{3,5,7\} & * & \left\{\frac{3}{5},\frac{4}{5},\frac{4}{5}\right\} \\ \left\{3,\frac{6}{5},\frac{4}{7}\right\} & \{1\} & \{4,6,9\} & * & \left\{\frac{16}{35},\frac{24}{25},\frac{18}{10}\right\} \\ \left\{\frac{1}{7},\frac{1}{5},\frac{1}{3}\right\} & \left\{\frac{1}{9},\frac{1}{6},\frac{1}{4}\right\} & \{1\} & * & * \\ * & * & * & \{1\} & \left\{\frac{1}{4},\frac{1}{4},\frac{1}{4}\right\} \\ \left\{\frac{5}{4},\frac{5}{4},\frac{5}{3}\right\} & \left\{\frac{10}{18},\frac{25}{24},\frac{35}{16}\right\} & * & \{4,4,4\} & \{1\} \end{bmatrix}$$

The missing element can be calculated using Eq. 8

$$\begin{aligned} \bar{b}_{e_{14}}^{\sigma(1)} &= (\bar{b}_{e_{15}}^{\sigma(3)} \times \bar{b}_{e_{54}}^{\sigma(3)}) = 16/5. \\ \bar{b}_{e_{14}}^{\sigma(2)} &= (\bar{b}_{e_{15}}^{\sigma(2)} \times \bar{b}_{e_{54}}^{\sigma(2)}) = 16/5. \\ \bar{b}_{e_{14}}^{\sigma(3)} &= (\bar{b}_{e_{15}}^{\sigma(1)} \times \bar{b}_{e_{54}}^{\sigma(1)}) = 12/5. \end{aligned}$$

Therefore

$$\bar{b}_{e_{14}} = \left\{ \bar{b}^{\sigma(1)}_{e_{14}}, \bar{b}^{\sigma(2)}_{e_{14}}, \bar{b}^{\sigma(3)}_{e_{14}} \right\}$$
$$= \left\{ \frac{16}{5}, \frac{16}{5}, \frac{12}{5} \right\}$$
$$\bar{b}^{\sigma(1)}_{e_{24}} = \left\{ (\bar{b}^{\sigma(3)}_{e_{21}} \times \bar{b}^{\sigma(3)}_{e_{14}}) \times (\bar{b}^{\sigma(3)}_{e_{25}} \times \bar{b}^{\sigma(3)}_{e_{54}}) \right\}^{1/2} = 3.142$$
$$\bar{b}^{\sigma(2)}_{e_{24}} = \left\{ (\bar{b}^{\sigma(2)}_{e_{21}} \times \bar{b}^{\sigma(2)}_{e_{14}}) \times (\bar{b}^{\sigma(2)}_{e_{25}} \times \bar{b}^{\sigma(2)}_{e_{54}}) \right\}^{1/2} = 3.84$$
$$\bar{b}^{\sigma(3)}_{e_{24}} = \left\{ (\bar{b}^{\sigma(1)}_{e_{21}} \times \bar{b}^{\sigma(1)}_{e_{14}}) \times (\bar{b}^{\sigma(1)}_{e_{25}} \times \bar{b}^{\sigma(1)}_{e_{54}}) \right\}^{1/2} = 4.189$$

Therefore,

$$\bar{b}_{e_{24}} = \left\{ \bar{b}^{\sigma(1)}_{e_{24}}, \bar{b}^{\sigma(2)}_{e_{24}}, \bar{b}^{\sigma(3)}_{e_{24}} \right\} = \{3.142, 3.84, 4.189\}$$
$$\bar{b}^{\sigma(1)}_{e_{34}} = \left\{ (\bar{b}^{\sigma(3)}_{e_{31}} \times \bar{b}^{\sigma(3)}_{e_{14}}) \times (\bar{b}^{\sigma(3)}_{e_{32}} \times \bar{b}^{\sigma(3)}_{e_{24}}) \right\}^{1/2} = 0.9153$$
$$\bar{b}^{\sigma(2)}_{e_{34}} = \left\{ (\bar{b}^{\sigma(2)}_{e_{31}} \times \bar{b}^{\sigma(2)}_{e_{14}}) \times (\bar{b}^{\sigma(2)}_{e_{32}} \times \bar{b}^{\sigma(2)}_{e_{24}}) \right\}^{1/2} = 0.64$$
$$\bar{b}^{\sigma(3)}_{e_{34}} = \left\{ (\bar{b}^{\sigma(1)}_{e_{31}} \times \bar{b}^{\sigma(1)}_{e_{14}}) \times (\bar{b}^{\sigma(1)}_{e_{32}} \times \bar{b}^{\sigma(1)}_{e_{24}}) \right\}^{1/2} = 0.399$$

Therefore,

$$\bar{b}_{e_{34}} = \left\{ \bar{b}^{\sigma(1)}_{e_{34}}, \bar{b}^{\sigma(2)}_{e_{34}}, \bar{b}^{\sigma(3)}_{e_{34}} \right\} = \{0.9153, 0.64, 0.399\}.$$
$$\bar{b}^{\sigma(1)}_{e_{35}} = \left\{ (\bar{b}^{\sigma(3)}_{e_{31}} \times \bar{b}^{\sigma(3)}_{e_{15}}) \times (\bar{b}^{\sigma(3)}_{e_{32}} \times \bar{b}^{\sigma(3)}_{e_{25}}) \times (\bar{b}^{\sigma(3)}_{e_{34}} \times \bar{b}^{\sigma(3)}_{e_{45}}) \right\}^{1/3} = 0.2287$$
$$\bar{b}^{\sigma(2)}_{e_{35}} = \left\{ (\bar{b}^{\sigma(2)}_{e_{31}} \times \bar{b}^{\sigma(2)}_{e_{15}}) \times (\bar{b}^{\sigma(2)}_{e_{32}} \times \bar{b}^{\sigma(2)}_{e_{25}}) \times (\bar{b}^{\sigma(2)}_{e_{34}} \times \bar{b}^{\sigma(2)}_{e_{45}}) \right\}^{1/3} = 0.16$$
$$\bar{b}^{\sigma(3)}_{e_{35}} = \left\{ (\bar{b}^{\sigma(1)}_{e_{31}} \times \bar{b}^{\sigma(1)}_{e_{15}}) \times (\bar{b}^{\sigma(1)}_{e_{32}} \times \bar{b}^{\sigma(1)}_{e_{25}}) \times (\bar{b}^{\sigma(1)}_{e_{34}} \times \bar{b}^{\sigma(1)}_{e_{45}}) \right\}^{1/3} = 0.09986$$

Therefore,

$$\bar{b}_{e_{35}} = \left\{ \bar{b}^{\sigma(1)}_{e_{35}}, \bar{b}^{\sigma(2)}_{e_{35}}, \bar{b}^{\sigma(3)}_{e_{35}} \right\} = \{0.2287, 0.16, 0.09986\}.$$

It is to note that the initial value of $\bar{b}_{e_{35}}$ not lies in between $1/9 - 9$ ratio scale, therefore error occurs. Thus the error is minimized using minimize the error, the above

optimization model (M) solved by MATLAB optimization tool box. The complete NH-MPR is given by

$$\bar{B}_R = \begin{bmatrix} \{1\} & \left\{\frac{7}{4}, \frac{5}{6}, \frac{1}{3}\right\} & \{3,5,7\} & \left\{\frac{16}{5}, \frac{16}{5}, \frac{12}{5}\right\} & \left\{\frac{3}{5}, \frac{4}{5}, \frac{4}{5}\right\} \\ \left\{3, \frac{6}{5}, \frac{4}{7}\right\} & \{1\} & \{4,6,9\} & \{3.142, 3.84, 4.189\} & \left\{\frac{16}{35}, \frac{24}{25}, \frac{18}{10}\right\} \\ \left\{\frac{1}{7}, \frac{1}{5}, \frac{1}{3}\right\} & \left\{\frac{1}{9}, \frac{1}{6}, \frac{1}{4}\right\} & \{1\} & \{0.915, 0.64, 0.399\} & \{0.229, 0.16, 0.111\} \\ \left\{\frac{5}{16}, \frac{5}{16}, \frac{5}{12}\right\} & \left\{\frac{1}{4.189}, \frac{1}{3.84}, \frac{1}{3.142}\right\} & \left\{\frac{1}{0.399}, \frac{1}{0.64}, \frac{1}{0.915}\right\} & \{1\} & \left\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right\} \\ \left\{\frac{5}{4}, \frac{5}{4}, \frac{5}{3}\right\} & \left\{\frac{10}{18}, \frac{25}{24}, \frac{35}{16}\right\} & \left\{\frac{1}{0.111}, \frac{1}{0.16}, \frac{1}{0.229}\right\} & \{4,4,4\} & \{1\} \end{bmatrix}$$

Similarly, using linear programming model (LP), the missing elements of incomplete NH-MPR $\bar{B}_R$ are obtained i.e., $\bar{b}_{14} = \{\frac{16}{5}, \frac{16}{5}, \frac{12}{5}\}$, $\bar{b}_{24} = \{3.142, 3.84, 4.189\}$, $\bar{b}_{34} = \{0.915, 0.64, 0.399\}$, $\bar{b}_{35} = \{0.229, 0.16, 0.111\}$. The above linear programming model (LP) is solved by LINGO software.

Once we have complete HMPR, our next goal is to check the consistency of the obtained relation. Therefore to validate the result, we have found the satisfaction index and acceptable consistency of HMPR which is given below.

4.1 Satisfaction index for HMPR

Zhu and Xu [27] gave a method to calculate the maximum satisfaction degree of HMPRs that can deal with hesitant judgement and produce a solution of priorities of the objectives of the decision makers. To accomplish the above task, Zhu and Xu [27] developed the following model

$$\begin{aligned} &\max \lambda \\ &\text{S.t.} \\ &t\lambda + \left(w_i - w_j \left(b^{\sigma(1)}_{e_{ij}} \; or \; \cdots \; or \; b^{|\sigma(l_{b_{e_{ij}}})|}_{e_{ij}} \right) \right) \leq t, \\ &t\lambda - \left(w_i - w_j \left(b^{\sigma(1)}_{e_{ij}} or \cdots or b^{|\sigma(l_{b_{e_{ij}}})|}_{e_{ij}} \right) \right) \leq t, \\ &i, j \in N, i < j \\ &\textstyle\sum_{i=1}^{n} w_i = 1, \; w_i \geq 0, \; i \in N \end{aligned} \tag{14}$$

where λ is used to measure the satisfaction degree of the decision maker to the solution and w_i, $w_i \geq 0$, $i \in N$ are the priority weights of the objectives and t is the deviation parameter such that if t is large enough, then $0 \leq \lambda \leq 1$; if t is small or the hesitant judgement given by the decision makers are very inconsistent and then $\lambda < 0$, though the value of priority weights remain same.

The satisfaction degree of the complete N-HMPR (for $t = 1$) of Example 1 is given by

$$
\begin{aligned}
&Max\ \lambda;\\
&subject\ to:\\
&\lambda + w_1 - \frac{7}{4}w_2 \le 1;\ \lambda + w_1 - \frac{5}{6}w_2 \le 1;\ \lambda + w_1 - \frac{1}{3}w_2 \le 1;\ \lambda + w_1 - 3w_3 \le 1;\\
&\lambda + w_1 - 5w_3 \le 1;\ \lambda + w_1 - 7w_3 \le 1;\ \lambda + w_1 - \frac{16}{5}w_4 \le 1;\ \lambda + w_1 - \frac{16}{5}w_4 \le 1;\\
&\lambda + w_1 - \frac{12}{5}w_4 \le 1;\ \lambda + w_1 - \frac{6}{10}w_5 \le 1;\ \lambda + w_1 - \frac{8}{10}w_5 \le 1;\ \lambda + w_1 - \frac{8}{10}w_5 \le 1;\\
&\lambda + w_2 - 4w_3 \le 1;\ \lambda + w_2 - 6w_3 \le 1;\ \lambda + w_2 - 9w_3 \le 1;\ \lambda + w_2 - 3.142w_4 \le 1;\\
&\lambda + w_2 - 3.84w_4 \le 1;\ \lambda + w_2 - 4.189w_4 \le 1;\ \lambda + w_2 - \frac{16}{35}w_5 \le 1;\ \lambda + w_2 - \frac{24}{25}w_5 \le 1;\\
&\lambda + w_2 - \frac{18}{10}w_5 \le 1;\ \lambda + w_3 - 0.915w_4 \le 1;\ \lambda + w_3 - 0.64w_4 \le 1;\ \lambda + w_3 - 0.399w_4 \le 1;\\
&\lambda + w_3 - 0.229w_5 \le 1;\ \lambda + w_3 - 0.16w_5 \le 1;\ \lambda + w_3 - 0.111w_5 \le 1;\ \lambda + w_4 - \frac{1}{4}w_5 \le 1;\\
&\lambda + w_4 - \frac{1}{4}w_5 \le 1;\ \lambda + w_4 - \frac{1}{4}w_5 \le 1;\ \lambda - w_1 + \frac{7}{4}w_2 \le 1;\ \lambda - w_1 + \frac{5}{6}w_2 \le 1;\\
&\lambda - w_1 + \frac{1}{3}w_2 \le 1;\ \lambda - w_1 + 3w_3 \le 1;\ \lambda - w_1 + 5w_3 \le 1;\ \lambda - w_1 + 7w_3 \le 1;\\
&\lambda - w_1 + \frac{16}{5}w_4 \le 1;\ \lambda - w_1 + \frac{16}{5}w_4 \le 1;\ \lambda - w_1 + \frac{12}{5}w_4 \le 1;\ \lambda - w_1 + \frac{6}{10}w_5 \le 1;\\
&\lambda - w_1 + \frac{8}{10}w_5 \le 1;\ \lambda - w_1 + \frac{8}{10}w_5 \le 1;\ \lambda - w_2 + 4w_3 \le 1;\ \lambda - w_2 + 6w_3 \le 1;\\
&\lambda - w_2 + 9w_3 \le 1;\ \lambda - w_2 + 3.142w_4 \le 1;\ \lambda - w_2 + 3.84w_4 \le 1;\ \lambda - w_2 + 4.189w_4 \le 1;\\
&\lambda - w_2 + \frac{16}{35}w_5 \le 1;\ \lambda - w_2 + \frac{24}{25}w_5 \le 1;\ \lambda - w_2 + \frac{18}{10}w_5 \le 1;\ \lambda - w_3 + 0.915w_4 \le 1;\\
&\lambda - w_3 + 0.64w_4 \le 1;\ \lambda - w_3 + 0.399w_4 \le 1;\ \lambda - w_3 + 0.229w_5 \le 1;\ \lambda - w_3 + 0.16w_5 \le 1;\\
&\lambda - w_3 + 0.111w_5 \le 1;\ \lambda - w_4 + \frac{1}{4}w_5 \le 1;\ \lambda - w_4 + \frac{1}{4}w_5 \le 1;\ \lambda - w_4 + \frac{1}{4}w_5 \le 1;\\
&w_1 + w_2 + w_3 + w_4 + w_5 = 1;\ w_1, w_2, w_3, w_4, w_5 \ge 0
\end{aligned}
$$

The satisfaction degree of complete HMPR is 0.8024 i.e., 80.24% and the weights vector are (0.2903, 0.2787, 0.0529, 0.11365, 0.2645). The above linear programming problem is solved by LINGO software. For different values of deviation parameter t, have different satisfaction degrees but same weights vector that is given in Table 1.

4.2 Acceptable consistency for HMPR

According to Zhang et al. [26], acceptable consistency of complete HMPR obtained from Example 1 is also checked. Using Eq. 4, complete HMPR $\bar{B}_R$ is split into three MPRs that is given as:

Table 1 Priorities obtained for different t

	$t = 1$	$t = 0.5$	$t = 0.3$
λ	0.8024	0.6052	0.3420
w_1	0.2903	0.2903	0.2903
w_2	0.2787	0.2787	0.2787
w_3	0.0529	0.0529	0.0529
w_4	0.11365	0.11365	0.11365
w_5	0.2645	0.2645	0.2645

$$
P^{(1)} = \begin{pmatrix}
1 & \frac{7}{4} & 3 & \frac{16}{5} & \frac{3}{5}\\
\frac{4}{7} & 1 & 4 & 3.142 & \frac{16}{35}\\
\frac{1}{3} & \frac{1}{4} & 1 & 0.915 & 0.229\\
\frac{5}{16} & \frac{1}{3.142} & \frac{1}{0.915} & 1 & \frac{1}{4}\\
\frac{5}{3} & \frac{35}{16} & \frac{1}{0.229} & 4 & 1
\end{pmatrix}_{5\times5}
\quad
P^{(2)} = \begin{pmatrix}
1 & \frac{5}{6} & 5 & \frac{16}{5} & \frac{4}{5}\\
\frac{6}{5} & 1 & 6 & 3.84 & \frac{24}{25}\\
\frac{1}{5} & \frac{1}{6} & 1 & 0.64 & 0.16\\
\frac{5}{16} & \frac{1}{3.84} & \frac{1}{0.64} & 1 & \frac{1}{4}\\
\frac{5}{4} & \frac{25}{24} & \frac{1}{0.16} & 4 & 1
\end{pmatrix}_{5\times5}
$$

$$
P^{(3)} = \begin{pmatrix}
1 & \frac{1}{3} & 7 & \frac{12}{5} & \frac{4}{5}\\
3 & 1 & 9 & 4.189 & \frac{18}{10}\\
\frac{1}{7} & \frac{1}{9} & 1 & 0.399 & 0.111\\
\frac{5}{12} & \frac{1}{4.189} & \frac{1}{0.399} & 1 & \frac{1}{4}\\
\frac{5}{4} & \frac{10}{18} & \frac{1}{0.111} & 4 & 1
\end{pmatrix}_{5\times5}
$$

Consistency ratio(CR) of MPRs $P^{(1)}$, $P^{(2)}$, $P^{(3)}$ is 0.0182, 0, 0.0169 respectively which are less than 0.1. Therefore the complete HMPR $\bar{B}_R$ is acceptable consistent.

For the equivalency of optimization model(M) and linear programming problem(LP) we have taken an another example.

Example 2 Consider a decision making problem with five sets of alternatives $x_i = \{x_1, x_2, x_3, x_4, x_5\}$. The decision maker judge these five alternatives by pairwise comparison and provides his/her judgement given in a matrix denoted by B_{R_1} given as follows:

$$
B_{R_1} = \begin{bmatrix}
\{1\} & \left\{2, \frac{4}{5}, \frac{1}{4}\right\} & \{2,4,6\} & * & \left\{\frac{2}{5}, \frac{2}{6}, \frac{2}{7}\right\}\\
\left\{4, \frac{5}{4}, \frac{1}{2}\right\} & \{1\} & \{3,5,8\} & * & \left\{\frac{1}{7}, \frac{5}{12}, \frac{8}{5}\right\}\\
\left\{\frac{1}{6}, \frac{1}{4}, \frac{1}{2}\right\} & \left\{\frac{1}{8}, \frac{1}{5}, \frac{1}{3}\right\} & \{1\} & * & *\\
* & * & * & \{1\} & \left\{\frac{1}{3}\right\}\\
\left\{\frac{7}{2}, \frac{6}{2}, \frac{5}{2}\right\} & \left\{\frac{5}{8}, \frac{12}{5}, 7\right\} & * & \{3\} & \{1\}
\end{bmatrix}
$$

Table 2 Estimation of missing elements

Missing values	Obtained from model (M)	Obtained from model (LP)
$(\bar{b}_{e_{14}}^{\sigma(1)}, \bar{b}_{e_{14}}^{\sigma(2)}, \bar{b}_{e_{14}}^{\sigma(3)})$	(0.857,1,1.2)	(0.857,1,1.2)
$(\bar{b}_{e_{24}}^{\sigma(1)}, \bar{b}_{e_{24}}^{\sigma(2)}, \bar{b}_{e_{24}}^{\sigma(3)})$	(1.697,1.25,1.212)	(1.697,1.25,1.212)
$(\bar{b}_{e_{34}}^{\sigma(1)}, \bar{b}_{e_{34}}^{\sigma(2)}, \bar{b}_{e_{34}}^{\sigma(3)})$	(0.492,0.25,0.174)	(0.492,0.25,0.174)
$(\bar{b}_{e_{35}}^{\sigma(1)}, \bar{b}_{e_{35}}^{\sigma(2)}, \bar{b}_{e_{35}}^{\sigma(3)})$	(0.164,0.111, 0.111)	(0.164,0.111, 0.111)

Table 3 Priorities obtained for different t

	$t = 1$	$t = 0.5$	$t = 0.3$
λ	0.8112654	0.3112654	0.1112654
w_1	0.2426	0.2426	0.2426
w_2	0.2157	0.2157	0.2157
w_3	0.0505	0.0505	0.0505
w_4	0.2383	0.2383	0.2383
w_5	0.2528	0.2528	0.2528

Using Algorithm 1, initial values of missing element of incomplete HMPR B_{R_1} are obtained. Initial values are $\bar{b}_{e_{14}} = \{0.857, 1, 1.2\}$, $\bar{b}_{e_{24}} = \{1.697, 1.25, 1.212\}$, $\bar{b}_{e_{34}} = \{0.492, 0.25, 0.174\}$, $\bar{b}_{e_{35}} = \{0.164, 0.0833, 0.058\}$.

It is to note that the initial value of $\bar{b}_{e_{35}}$ not lies in between $1/9 - 9$ ratio scale, therefore error occurs. Thus the error is minimized using minimize the error, the above optimization model (M) solved by MATLAB optimization tool box. Also to calculate the missing value we have solve by an linear programming problem give the same result with optimization model(M) given in the Table 2.

The satisfaction index of complete HMPR of Example 2 is 0.8112654 i.e., 81.12% and the weights vector are given in Table 3. The above linear programming problem is solved by LINGO software. For different values of deviation parameter t, have different satisfaction degrees but weights vector are same.

Also we have the split the complete hesitant multiplicative preference relation B_{R_1} into three multiplicative preference relation using Eq. 4. The consistency ratio of three MPRs are 0.0629, 0.0022, 0.0615 which are less than 0.1. Therefore, B_{R_1} is acceptable consistent.

5 Conclusion

In this paper, we have defined a new multiplicative transitive property of HMPR. We have developed an algorithm to construct a complete HMPRs from an incomplete one. It may be the case that the initial values obtained from Algorithm 1 does not lies in the ratio scale $1/9 - 9$. We have developed an optimization model to minimize the error. Also, A linear programming problem is developed to find out the missing element of HMPR. Satisfaction degree and acceptable consistency of complete HMPR is also checked. We have presented two example to illustrates the above method. The present work can be extended to intuitionistic hesitant multiplicative preference relation, hesitant linguistic preference relation and intuitionistic hesitant linguistic preference relation and their application in decision-making problem.

References

1. Atanassov, K.: Intuitionistic fuzzy sets. Fuzzy Sets Syst. **20**(1), 87–96 (1986)
2. Dong, Y., Xu, Y., Li, H.: On consistency measures of linguistic preference relations. Eur. J. Oper. Res. **189**(2), 430–444 (2008)
3. Herrera-Viedma, E., Alonso, S., Chiclana, F., Herrera, F.: A Consensus Model for group decision making with incomplete fuzzy preference relations. IEEE T Fuzzy Syst. **15**, 863–877 (2007)
4. Orlovsky, S.A.: Decision-making with a fuzzy preference relation. Fuzzy sets syst. **1**(3), 155–167 (1978)
5. Saaty, T.L.: A scaling method for priorities in hierarchical structures. J. Math. Psychol. **15**, 234–281 (1977)
6. Saaty, T.L.: The Analytical Hierarchy Process. MC Graw-Hill, New York (1980)
7. Tanino, T.: Fuzzy preference ordering in group decision making. Fuzzy Sets Syst. **12**(2), 117–131 (1984)
8. Torra, V.: Hesitant fuzzy sets. Int. J. Intell. Syst. **25**(6), 529539 (2010)
9. Xia, M.M., Xu, Z.S., Chen, J.: Algorithms for improving consistency or consensus of reciprocal [0,1]-valued preference relations. Fuzzy Sets Syst. **216**, 108–133 (2013)
10. Xia, M.M., Xu, Z.S., Liao, H.C.: Preference relations based on intuitionistic multiplicative information. IEEE Trans. Fuzzy Syst. **21**(1), 113–133 (2013)
11. Xia, M.M., Xu, Z.S.: Managing hesitant information in GDM problems under fuzzy and multiplicative preference relations. Int. J. Uncertain. Fuzziness Knowl. Based Syst. **21**, 865–897 (2013)
12. Xu, Y., Chen, L., Rodrguez, R.M., Herrera, F., Wang, H.: Deriving the priority weights from incomplete hesitant fuzzy preference relations in group decision making. Knowl.-Based Syst. **99**, 71–78 (2016)
13. Xu, Z.S.: A method for priorities of triangular fuzzy number complementaryjudgment matrices. Fuzzy Syst. Math. **16**, 47–50 (2002)
14. Xu, Z.S., Cai, X.Q.: Incomplete interval valued intuitionistic fuzzy preference relations. Int. J. Gen. Syst. **38**, 871–886 (2009)
15. Xu, Z.S.: Goal programming models for obtaining the priority vector of incom-plete fuzzy preference relation. Int. J. Approx. Reason. **36**(3), 261270 (2004)
16. Xu, Z.S.: Intuitionistic Fuzzy Preference Modeling and Interactive Decision Making. Springer, Berlin (2013)
17. Xu, Z.S.: Intuitionistic preference relations and their application in group decision making. Inf. Sci. **177**(11), 2363–2379 (2007)
18. Xu, Z.S.: On consistency of the weighted geometric mean complex judgmentmatrix in AHP. Eur. J. Oper. Res. **126**(3), 683–687 (2000)
19. Xu, Z.S., Wei, C.P.: A consistency improving method in the analytic hierarchy process. Eur. J. Oper. Res. **116**(2), 443–449 (1999)
20. Xu, Z.S., Xia, M.M.: Distance and similarity measures for hesitant fuzzy sets. Inf. Sci. **181**(11), 2128–2138 (2011)
21. Xu, Z.S., Yager, R.R.: Intuitionistic and interval-valued intutionistic fuzzy preference relations and their measures of similarity for the evaluation of agreement within a group. Fuzzy Optim. Decis. Mak. **8**, 123–139 (2009)
22. Yu, D.J., Merig, J.M., Zhou, L.G.: Interval valued multiplicative intuitionistic fuzzy preference relations. Int. J. Fuzzy Syst. **15**(4), 412–422 (2013)
23. Zhang, Z., Guo, C.: Fusion of heterogeneous incomplete hesitant preference relations in group decision making. Int. J. Comput. Intell. Syst. **9**(2), 245–262 (2016)

24. Zhang, Z., Kou, X., Dong, Q.: Additive consistency analysis and improvement for hesitant fuzzy preference relations. Expert Syst. Appl. **98**, 118–128 (2018)
25. Zhang, Z., Kou, X., Yu, W., Guo, C.: On priority weights and consistency for incomplete hesitant fuzzy preference relations. Knowl.-Based Syst. **000**, 1–12 (2017)
26. Zhang, Z., Wu, C.: Deriving the priority weights from hesitant multiplicative preference relations in group decision making. Appl. Soft Comput. **25**, 107–117 (2014)
27. Zhu, B., Xu, Z.: Analytic hierarchy process hesitant group decision making. Eur. J. Oper. Res. **239**(3), 794–801 (2014)
28. Zhu, B., Xu, Z.: Regression methods for hesitant fuzzy preference relations. Technol. Econ. Dev. Econ. **19**(Suppl. 1), S214–S227 (2013)

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

Analog Integrated Circuits and Signal Processing
https://doi.org/10.1007/s10470-019-01410-4

MIXED SIGNAL LETTER

Lossless grounded admittance simulator using OTRA

Bal Chand Nagar[1] · **Sajal K. Paul**[2]

Received: 1 October 2018 / Revised: 26 December 2018 / Accepted: 16 January 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract
In this paper, a generalized admittance simulator is presented. The proposed circuit uses two operational transresistance amplifiers (OTRAs), one voltage buffer and three passive components. The proposed generalized admittance simulator can realize three simulators namely lossless grounded positive inductance simulator, frequency dependent negative resistance (FDNR) simulator and capacitance simulator/multiplier without any matching constraint. The non-ideality analysis of the circuit is also given. The workability of the proposed admittance simulator is shown through the implementation of second order band pass filter using inductance simulator, second order low pass filter using FDNR and variable capacitance low pass filter using capacitance multiplier. PSPICE simulation results are included to verify theory.

Keywords Inductance simulator · FDNR simulator · Capacitance simulator · Operational transresistance amplifier · Integrated circuit

1 Introduction

Recently, a lot of effort has been devoted by researchers to the design of the operational transresistance amplifier (OTRA) and circuit based on it. The ideal OTRA is insensitive to parasitic capacitances due to internally grounded input terminals. It can provide a high gain, high bandwidth and slew rate unlike traditional operational amplifier [1–6]. Literature survey reveals that an OTRA is used for a number of signal generation and processing applications such as active simulators [7–16], oscillators [17, 18], filters [19], multivibrators [20], schmitt trigger [21], PID controller [22] and rectifiers [23] etc.

Active simulators such as inductors, frequency dependent negative resistors (FDNRs) and capacitance simulators/multipliers are useful elements for designing a wide range of circuits such as active filters, oscillators, phase shifters, cancellation of parasitic element and impedance matching. During the last few years many active simulators are proposed using OTRAs [7–16]. Most of these are inductance simulators. Only few simulators [11, 12] are multi-functional. All these simulators use matching constraints and more than three passive components.

In this paper a new lossless grounded general multi-functional simulator is presented. This circuit can be used to implement lossless grounded inductance, FDNR and capacitance simulators without any matching constraint with minimum passive components. The proposed admittance simulator uses only two OTRAs, one buffer and three passive components (either two resistors and one capacitor or two capacitors and one resistor). The proposed admittance simulator is simulated in PSPICE using CMOS based OTRA. Various application examples are included to confirm the theoretical proposition.

The organization of the paper starts with this Sect. 1, where introduction is given. The description of active building block and proposed lossless grounded admittance simulator as a lossless grounded positive inductor, FDNR and capacitance simulators are presented in Sect. 2. The effects of non-ideality of OTRA on admittance simulator are discussed in Sect. 3. The simulation results with an application example for each simulator are presented in Sect. 4 followed by Sect. 5 where experimental verification

✉ Sajal K. Paul
sajalkpaul@rediffmail.com

Bal Chand Nagar
balchandnagar@gmail.com

[1] Department of Electronics and Communication Engineering, National Institute of Technology Patna, Ashok Rajpath, Patna, Bihar 800005, India

[2] Department of Electronics Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, Jharkhand 826004, India

Published online: 24 January 2019

for use of inductor simulator is given. Comparative study of proposed simulator with other OTRA based reported circuits are presented in Sect. 6 followed by conclusion in Sect. 7.

2 Proposed circuit

The OTRA is a three-port building block and the symbol is given in Fig. 1(a), and its terminal equations can be given as

$$\begin{bmatrix} V_p \\ V_n \\ V_0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ R_m & -R_m & 0 \end{bmatrix} \begin{bmatrix} I_p \\ I_n \\ I_0 \end{bmatrix} \quad (1)$$

where R_m is dc transresistance gain of OTRA. Ideally, the R_m, approaches infinity; hence the use of OTRA in the external negative feedback will force the input currents, I_p and I_n, to be equal. Among the MOS based high performance OTRA realizations available in literature, the structure used for simulation is shown in Fig. 1(b) [3, 4]. On the other hand OTRA can be implemented using commercially available IC (AD844AN) [5, 6] as shown in Fig. 1(c).

The proposed simulator topology using two OTRAs, one voltage follower (VF) and three passive components is shown in Fig. 2. The routine analysis of the circuit is given below.

The input currents of OTRA1 are given by

$$I_{n1} = I_{in} \quad (2)$$

$$I_{p1} = V_0 Y_2 \quad (3)$$

In an OTRA ideally, the R_m approaches infinity which forces input currents I_p and $I_n$to be equal. Hence we can write

$$I_{p1} = I_{n1} \quad (4)$$

$$I_{in} = V_0 Y_2 \quad (5)$$

Similarly, from OTRA2, we can get

$$V_{in} Y_1 = V_0 Y_3 \quad (6)$$

From (5) and (6), we get the expression for the input admittance as

$$Y_{in} = \frac{I_{in}}{V_{in}} = \frac{Y_1 Y_2}{Y_3} \quad (7)$$

It is shown in the following sections that the proper choice of Y_1, Y_2 and Y_3 results lossless inductor, FDNR and grounded capacitor simulators. It is also noting that floating resistors can easily be implemented using two NMOS transistors working in the linear region with complete nonlinearity cancellation, because OTRA input terminals are virtually grounded [3]. It is also seen that each proposed circuit employs floating capacitor, but they can be easily realized with a current IC process that offers double poly (poly1-poly2) or metal–insulator-metal (MIM) capacitor or metal-oxide-metal (MOM) capacitor [24, 25].

2.1 Lossless grounded positive inductor

The proposed lossless grounded positive inductor simulator is obtained by selecting Y_1 and Y_2 as resistors and Y_3 as a capacitor as shown in Fig. 3(a) which results the following expression for the input admittance:

$$Y_{in} = \frac{1}{sC_1R_1R_2} \quad (8)$$

Thus, the circuit realizes a lossless grounded inductor simulator whose value is given by

$$L_{eq} = C_1R_1R_2 \quad (9)$$

The sensitivity of L_{eq} to the variation in passive components is given by

$$S_{C_1}^{L_{eq}} = S_{R_1}^{L_{eq}} = S_{R_2}^{L_{eq}} = 1 \quad (10)$$

It shows that the sensitivity of the proposed circuit is low. It is also observed that the value of inductance can easily be tuned with R_1 and/or R_2.

2.2 Lossless grounded positive FDNR

The proposed lossless grounded positive FDNR simulator is obtained by selecting Y_1 and Y_2 as capacitors and Y_3 as a resistor and the corresponding circuit is shown in Fig. 3(b). The input admittance Y_{in} of the circuit is given by the following equation:

$$Y_{in} = s^2C_1C_2R_1 \quad (11)$$

Thus, the circuit realizes a lossless grounded FDNR whose value is given by

$$D_{eq} = C_1C_2R_1 \quad (12)$$

The sensitivity of D_{eq} to the variation in passive componnts is given by

$$S_{C_1}^{D_{eq}} = S_{C_2}^{D_{eq}} = S_{R_1}^{D_{eq}} = 1 \quad (13)$$

It shows that the sensitivity of the proposed circuit is low, and the value of D_{eq} can easily be controlled by R_1.

2.3 Lossless grounded positive capacitance simulator/multiplier

The proposed lossless grounded capacitance simulator is obtained by two ways first by choosing Y_1 as a capacitor and Y_2 and Y_3 as resistors and the second way is by

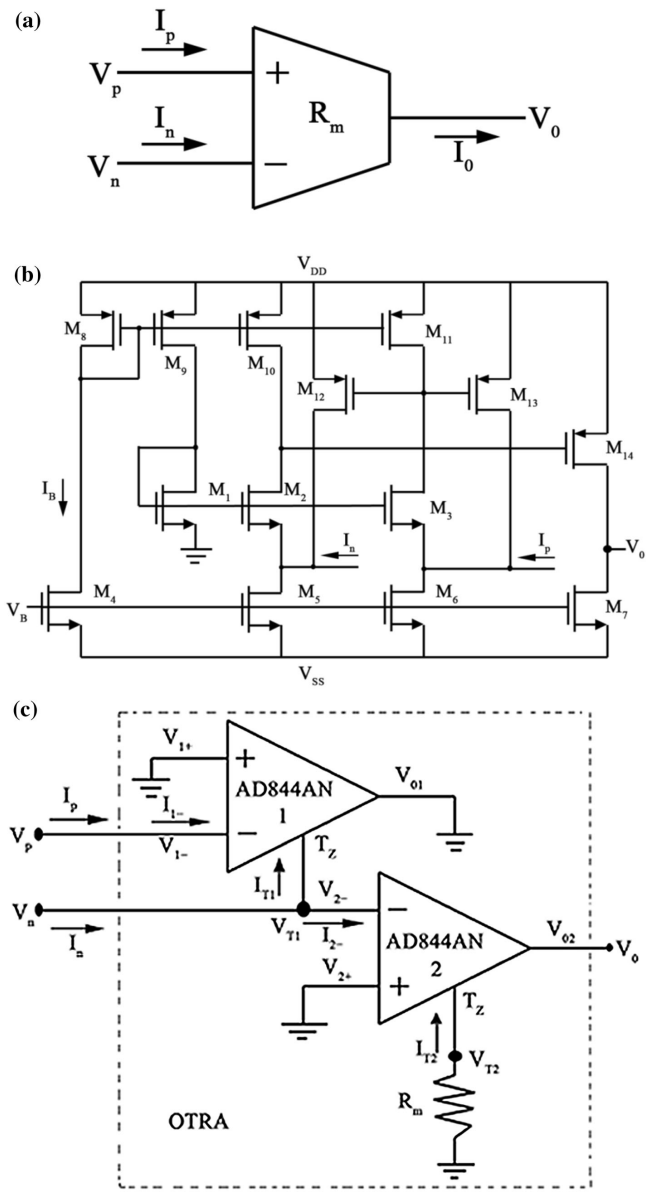


Fig. 1 OTRA: **a** ideal model **b** internal structure using CMOS **c** implementation using CFOA IC (AD844AN) [5]

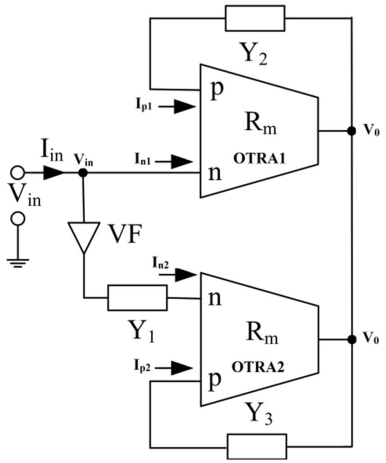


Fig. 2 The proposed lossless grounded admittance simulator

choosing Y_2 as a capacitor and Y_1 and Y_3 as resistors and the corresponding circuits are shown in Fig. 3(c) and (d) respectively. The input admittance Y_{in} of the circuit of Fig. 3(c) results:

$$Y_{in1} = sC_1\frac{R_3}{R_2} \tag{14}$$

Thus, the circuit realizes a lossless grounded capacitance simulator/multiplier whose value is given by

$$C_{eq1} = C_1\frac{R_3}{R_2} \tag{15}$$

The sensitivity of C_{eq1} to the variation in passive components is given by

$$S_{C_1}^{C_{eq1}} = S_{R_3}^{C_{eq1}} = 1, \quad S_{R_2}^{C_{eq1}} = -1 \tag{16}$$

Similarly the capacitance obtained for the circuit of Fig. 3(d) is

$$C_{eq2} = C_2\frac{R_3}{R_1} \tag{17}$$

and the corresponding sensitivities are

$$S_{C_2}^{C_{eq2}} = S_{R_3}^{C_{eq2}} = 1, \quad S_{R_1}^{C_{eq2}} = -1 \tag{18}$$

It is clear that the multiplication factor of capacitor simulator can easily be varied by resistances and also sensitivity is found to be low whose magnitude is unity.

3 Non-ideality analysis

In this section, the effect of the non-ideal transresistance gain (R_m) in the OTRA on the proposed admittance simulator is discussed. Till now, analysis has been done assuming ideal OTRA (infinite and frequency independent R_m). However, practically the R_m is a frequency dependent finite value. Considering the single pole model, the R_m can be expressed as:

$$R_m(s) = \left(\frac{R_{m0}}{1+\frac{s}{\omega_0}}\right) \tag{19}$$

where R_{m0} is the dc value of transresistance gain. At high frequency, the $R_m(s)$ reduce to

$$R_m(s) \approx \left(\frac{1}{sC_p}\right), \text{ where } C_p = \left(\frac{1}{R_{m0}\omega_0}\right) = \text{parasitic capacitance.} \tag{20}$$

The analysis of the proposed admittance simulator circuit (Fig. 2) taking non-ideality effect into account yields a modified expression for $Y_{in}(s)$ as,

$$Y_{in} = \frac{Y_1Y_2}{Y_3}\frac{\varepsilon_{uc1}(s)}{\varepsilon_{uc2}(s)}, \tag{21}$$

where $\varepsilon_{uc1}(s)$ and $\varepsilon_{uc2}(s)$ are the uncompensated error function and are expressed as

$$\varepsilon_{uc1}(s) = \left(1-\frac{sC_{p1}}{Y_2}\right), \text{ and } \varepsilon_{uc2}(s) = \left(1-\frac{sC_{p2}}{Y_3}\right), \tag{22}$$

where C_{p1} and C_{p2} are C_p of two OTRAs. The errors $\varepsilon_{uc1}(s)$ and $\varepsilon_{uc2}(s)$ appears in the admittance simulator topology of Fig. 2 at high frequency may be eliminated (compensated) by employing a modified topology. High frequency passive compensated topology of the simulated admittance with two additional elements Y_{p1} and Y_{p2} is shown in Fig. 4. Routine analysis results

$$Y_{in} = \frac{Y_1Y_2}{Y_3}\frac{\varepsilon_{c1}(s)}{\varepsilon_{c2}(s)}, \tag{23}$$

where $\varepsilon_{c1}(s)$ and $\varepsilon_{c2}(s)$ are the compensated error function and are expressed as

$$\varepsilon_{c1}(s) = \left[1+\left(\frac{Y_{p1}-sC_{p1}}{Y_2}\right)\right], \text{ and } \varepsilon_{c2}(s) = \left[1+\left(\frac{Y_{p2}-sC_{p2}}{Y_3}\right)\right], \tag{24}$$

The inspection of (24) suggests that $Y_{p1} = sC_{p1}$ and $Y_{p2} = sC_{p2}$ reduces $\varepsilon_{c1}(s)$ and $\varepsilon_{c2}(s)$ to 1, and hence (23) reduces to (7). It implies that the effect of single pole model of R_m can thus be eliminated by connecting a capacitors having value C_{p1} and C_{p2} in place of Y_{p1} and Y_{p2}. The above result is applicable for all the realize simulators namely lossless grounded positive inductance simulator, FDNR simulator and capacitance simulator by proper choice of Y_1, Y_2 and Y_3.

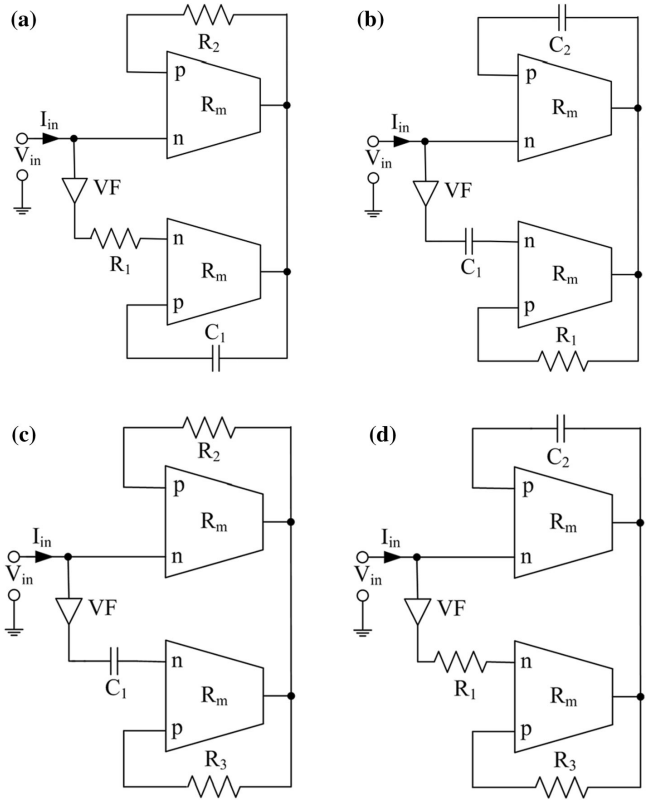


Fig. 3 Proposed lossless grounded simulator: **a** inductance; **b** FDNR; **c** capacitance multiplier-I; **d** capacitance multiplier-II

4 Simulation results and application examples

The workability of the proposed admittance simulator is simulated using PSPICE simulator in 180 nm CMOS technology. The aspect ratios of PMOS and NMOS transistors are given in Table 1 and supply voltages used are $V_{DD} = -V_{SS} = 0.9$ V and $V_B = -0.3$ V. There are many MOS based voltage followers (VF) available in literature. In Ref. [26], the MOS based VF is used for simulation with supply voltages $V_{DD} = -V_{SS} = 0.9$ V, the bias voltage $V_B = -0.3$ V and the aspect ratios of the M_1 and M_2 transistors as 260 µm/0.52 µm and 169 µm/0.52 µm respectively.

To design an inductance of 1 mH the passive components of $R_1 = R_2 = 1$ kΩ and $C_1 = 1$ nF are used. The simulated impedance magnitude for 1 mH inductance is shown in Fig. 5. It is found that inductance value remains within ± 10% in the frequency range of 1 kHz–9.8 MHz. Similarly, to implement an FDNR of 1 fFs (femto-farad-second) the passive elements of $C_1 = C_2 = 1$ nF and $R_1 = 1$ kΩ are used. The simulated magnitude for 1 fFs FDNR is shown in Fig. 6. It is observed from Fig. 6 that the FDNR operates properly up to 12 MHz. Similarly to implement a capacitance simulator/multiplier of 1 nF the passive elements of $C_1 = 1$ pF, $R_2 = 100$ kΩ and $R_3 = 100$ Ω are used. The simulated magnitude for 1 nF capacitance is shown in Fig. 7. It is observed from Fig. 7

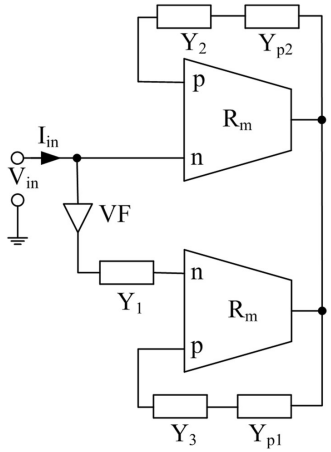


Fig. 4 High frequency compensation

Table 1 Aspect ratios of MOS transistors of the OTRA in Fig. 1(b)

Transistor	W (μm)/L (μm)
M1–M3	36/0.9
M4	3.6/0.9
M5, M6	10.8/0.9
M7	3.6/0.9
M8–M11	18/0.9
M12, M13	36/0.9
M14	18/0.18

that the capacitance simulator operates properly up to 9 MHz. The total power consumption of the proposed circuit is found to be 833 μW.

4.1 Second order band pass filter (BPF) using simulated inductor

A BPF shown in Fig. 8(a) is implemented using proposed inductor simulator. The transfer function computed for BPF is given as

$$T(s) = \frac{V_0(s)}{V_i(s)} = \frac{\frac{s}{C_B R_B}}{s^2 + \frac{s}{C_B R_B} + \frac{1}{L_{eq} C_B}} \tag{25}$$

From (25) the pole frequency (ω_0), bandwidth (ω_0/Q) and quality factor (Q) can be expressed as

$$\omega_0 = \frac{1}{\sqrt{L_{eq} C_B}}, \frac{\omega_0}{Q} = \frac{1}{C_B R_B}, \text{ and } Q = R_B \sqrt{\frac{C_B}{L_{eq}}} \tag{26}$$

The performance of the BPF of Fig. 8(a) is verified using the inductor simulator topology of Fig. 3(a) for a design having a pole frequency of 159 kHz for Q = 1 with R_B = 1 kΩ, C_B = 1 nF and L_{eq} = 1 mH (R_1 = R_2 = 1 kΩ and C_1 = 1 nF). The frequency response of the BPF is shown in Fig. 8(b). Simulated value of pole frequency is found to be 158.86 kHz which is close to the theoretical value.

4.2 Second order low pass filter (LPF) using simulated FDNR

A second order LPF shown in Fig. 9(a) is constructed using proposed FDNR simulator. The transfer function obtained for LPF is given as

$$T(s) = \frac{V_0(s)}{V_i(s)} = \frac{\frac{1}{D_{eq} R_L}}{s^2 + \frac{s}{C_L R_L} + \frac{1}{D_{eq} R_L}} \tag{27}$$

From (27) the pole frequency (ω_0), bandwidth (ω_0/Q) and quality factor (Q) can be expressed as

$$\omega_0 = \frac{1}{\sqrt{D_{eq} R_L}}, \frac{\omega_0}{Q} = \frac{1}{C_L R_L}, \text{ and } Q = C_L \sqrt{\frac{R_L}{D_{eq}}} \tag{28}$$

Fig. 5 Inductance response for L_{eq} = 1 mH

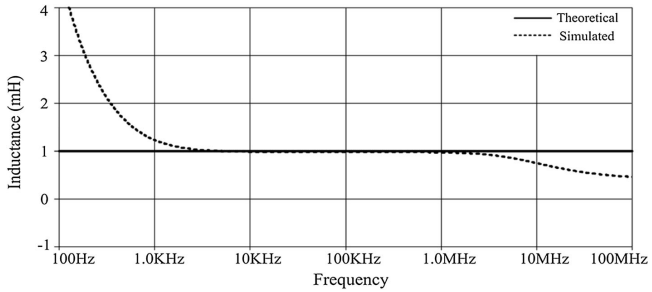


Fig. 6 FDNR response for D_{eq} = 1 fFs

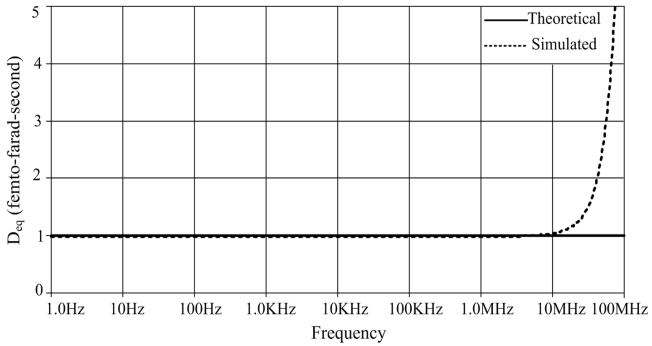


Fig. 7 Capacitance multiplier response for C_{eq1} = 1 nF

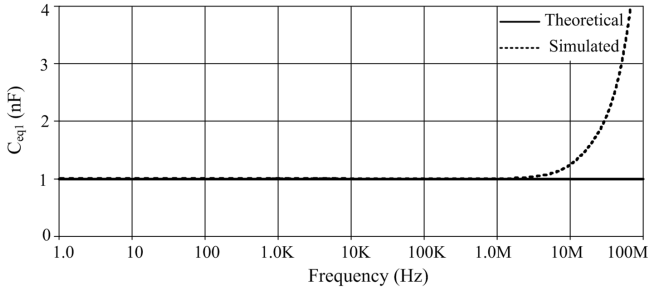
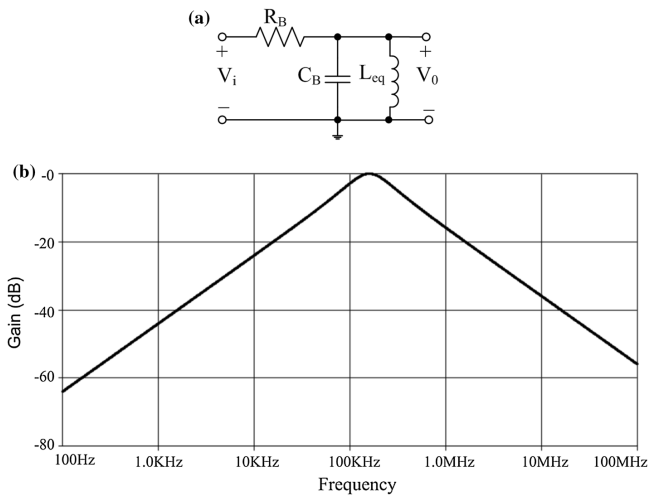


Fig. 8 BPF: **a** using inductance simulator; **b** frequency response



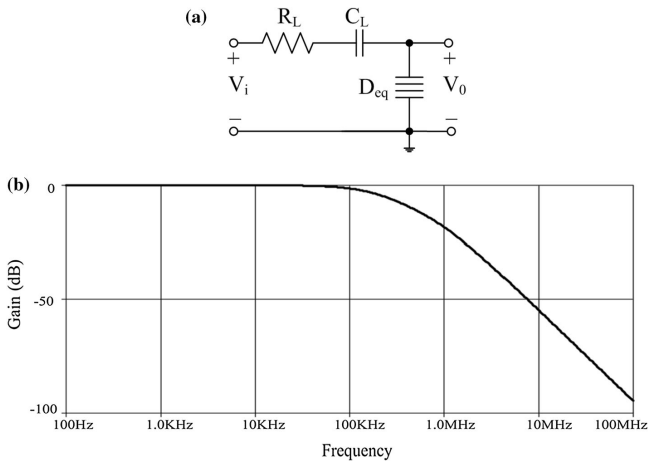


Fig. 9 LPF: **a** using FDNR simulator; **b** frequency response

where $D_{eq} = C_1C_2R_1$.

The performance of the LPF of Fig. 9(a) of is verified using the FDNR topology of Fig. 3(b) for a design having a pole frequency of 159 kHz for Q = 1 with R_L = 1 kΩ, C_L = 1 nF and D_{eq} = 1 fFs (C_1 = C_2 = 1 nF and R_1 = 1 kΩ). The frequency response of the LPF is shown in Fig. 9(b). Simulated value of pole frequency is 159.2 kHz, which agrees closely with theoretical value.

4.3 First order low pass filter (LPF) using simulated capacitance simulator

A first order LPF shown in Fig. 10(a) is constructed using proposed capacitance simulator. The transfer function obtained for LPF is given as

$$T(s) = \frac{V_0(s)}{V_i(s)} = \frac{1}{1 + sC_{eq1}R_L} \tag{29}$$

From (29) the pole frequency (ω_0) can be expressed as

$$\omega_0 = \frac{1}{C_{eq1}R_L}, \text{ where } C_{eq1} = C_1\frac{R_3}{R_2} \tag{30}$$

The performance of the LPF of Fig. 10(a) is verified using the capacitance simulator/multiplier-I topology of Fig. 3(c) for a design having a pole frequency of 159 kHz with R_L = 1 kΩ, C_{eq1} = 1 nF (C_1 = 1 pF, R_2 = 100 kΩ and R_3 = 100 Ω). The frequency response of the LPF is shown in Fig. 10(b). Simulated value of pole frequency (158.92 kHz) agrees closely with the theoretical value.

5 Experimental verification

The functionality of the proposed topology is also verified experimentally for the inductance simulator with an application example of LC oscillator. An LC oscillator can be realized using the proposed lossless grounded inductor simulator (L_{eq}) of Fig. 3(a), which is shown in Fig. 11. The characteristics equation obtained for the LC oscillator is obtained as

$$s^2 + \frac{s}{C_{O1}R_{O1}R_{O2}}[R_{O1} + R_{O2} - R_{O3}] + \frac{1}{L_{eq}C_{O1}} = 0 \tag{31}$$

For which the condition of oscillation (CO) and frequency of oscillation (FO) are obtained as

$$\text{CO: } R_{O1} + R_{O2} = R_{O3} \tag{32}$$

$$\text{FO: } \omega_0 = \frac{1}{\sqrt{L_{eq}C_{O1}}} \tag{33}$$

The experimental verification of the LC oscillator of Fig. 11 is done using the OTRA realization of commercially available CFOA IC AD844AN [5, 6] of Fig. 1(c). The voltage follower (VF) has been implemented using a CFOA. The DC bias supply of the CFOA IC AD844AN is chosen as ± 5 V and the component values are chosen for the circuit of Fig. 11 as R_{O1} = 1 kΩ, R_{O2} = 1 kΩ, R_{O3} = 2 kΩ, C_{O1} = 1 nF and L_{eq} = 1 mH (R_1 = R_2 = 1 kΩ and C_1 = 1 nF). Theoretically calculated frequency for chosen component values is 159 kHz. The experimental waveform obtained on the oscilloscope is shown in Fig. 12. The observed FO is found to be 158.20 kHz, which is in close agreement to theoretically calculated value of 159 kHz.

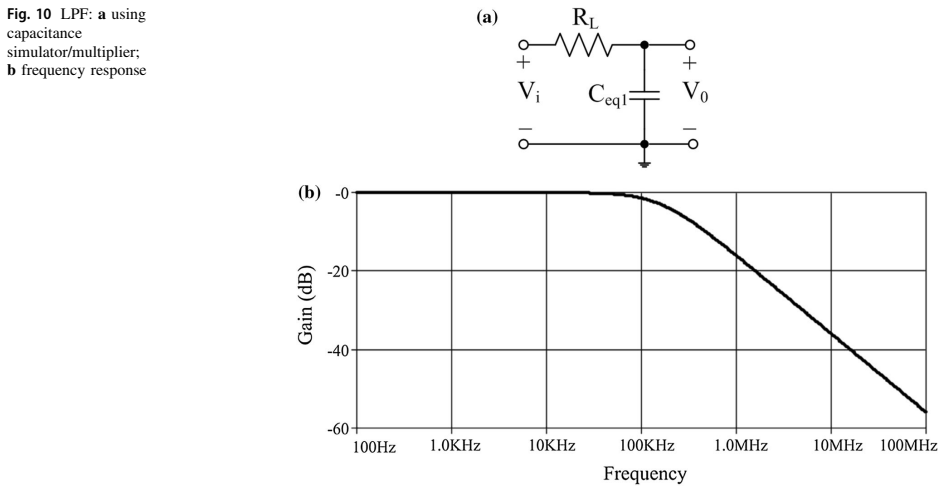


Fig. 10 LPF: **a** using capacitance simulator/multiplier; **b** frequency response

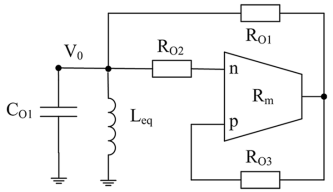


Fig. 11 LC oscillator

6 Comparison

Many grounded active inductors, FDNRs and Capacitance simulators using OTRA are available in literature [7–16]. The comparisons of these simulators are given in Table 2.

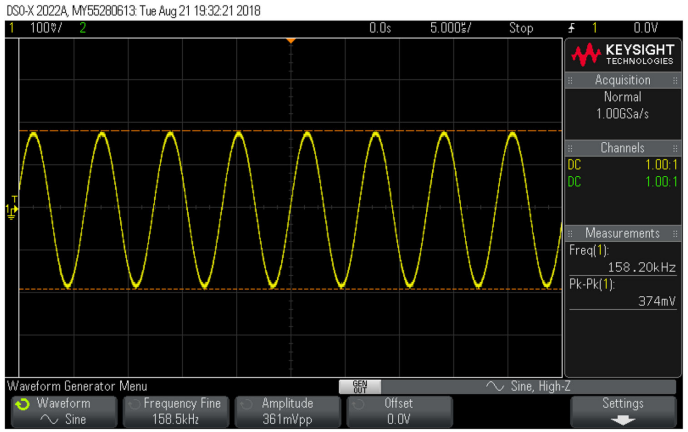


Fig. 12 Experimental result of the LC oscillator

Table 2 A comparison of lossless grounded simulators using OTRA

References	No. of ABBs	No. of Passive components (R/C)	Realization (L_{eq}, FDNR, C_{eq})	Passive component matching required
Ref. [7] (Fig. 6)	OTRA = 2	6 (5/1)	L_{eq}	Yes
Ref. [8] (Fig. 2c)	OTRA = 2	6 (5/1)	L_{eq}	Yes
Ref. [9]	OTRA = 1	6 (5/1)	$-L_{eq}$	Yes
Ref. [10]	OTRA = 2	6 (5/1)	L_{eq}	Yes
Ref. [11]	OTRA = 1, VF = 1	4 (2/2)	L_{eq}	Yes
		4 (2/2)	FDNR	Yes
Ref. [12] (Fig. 4)	OTRA = 2, VF = 1	6 (5/1)	L_{eq}	Yes
		6 (4/2)	FDNR	Yes
		6 (5/1)	C_{eq}	Yes
Ref. [13]	OTRA = 1	5 (3/2)	L_{eq}	Yes
Ref. [14]	OTRA = 1	5 (4/1)	$-L_{eq}$	Yes
Ref. [15]	OTRA = 1	4 (3/1)	$-L_{eq}$	Yes
Ref. [16]	OTRA = 1	5(2/3)	FDNR	Yes
Proposed	OTRA = 2, VF = 1	3 (2/1)	L_{eq}	No
		3 (1/2)	FDNR	No
		3 (2/1)	C_{eq}	No

The inspection of Table 2 reveals that most of the literatures based on OTRA simulate only inductance [7–10, 13–15]. The work in Ref. 11, uses one less number of OTRA in count in comparison to the proposed one, however it simulates only inductance (L_{eq}) and FDNR and uses one excess passive component and also requires passive components matching in contrast to the proposed one. Although, the work in Ref. [12], uses same number of active building blocks (ABBs), it uses three excess passive components for realization of L_{eq}, FDNR and C_{eq} in contrast to the proposed one. The work in Ref. 16 implements only FDNR and requires passive components matching; moreover it uses five passive components in contrast to three in the proposed circuits.

7 Conclusion

In this paper, a generalized admittance simulator is proposed. This can be used to implement lossless grounded inductor, FDNR and capacitance simulators with minimum passive components and without any matching constraint. The proposed admittance simulator employs two OTRAs, one buffer and three passive components. To demonstrate the practical use of the proposed circuit, some application examples (second order band pass filter using inductance simulator, second order LPF using FDNR and first order LPF using capacitance multiplier) have been included. It is also found that the non-linearity error may be eliminated at high frequency by passive compensation. The propose circuit offers some advantageous features such as: (i) no matching constraint, (ii) minimum passive components, (iii) low passive sensitivity, and (iv) low power consumption. Simulation and experimental results verify the theoretical propositions.

References

1. Chen, J. J., Tsao, H. W., & Chen, C. C. (1992). Operational transresistance amplifier using CMOS technology. *Electronics Letters, 28*(22), 2087–2088.
2. Chen, J. J., Tsao, H. W., Liu, S. I., & Chiu, W. (1995). Parasitic-capacitance-insensitive current-mode filters using operational transresistance amplifiers. *IEE Proceedings, Circuits Devices Systems, 142*(3), 186–192.
3. Salama, K. N., & Soliman, A. M. (1999). CMOS operational transresistance amplifier for analog signal processing. *Microelectronics Journal, 30*(3), 235–245.
4. Mostafa, H., & Soliman, A. M. (2006). A modified CMOS realization of the operational transresistance amplifier (OTRA). *Frequenz, 60*(3–4), 70–76.
5. Hou, C. L., Chien, H. C., & Lo, Y. K. (2005). Squarewave generators employing OTRAs. *IEE Proceedings, Circuits Devices Systems, 152*(6), 718–722.
6. Analog Devices Inc. (1999). *Linear products data book*. Norwood, MA: Analog Devices Inc.
7. Salama, K. N., & Soliman, A. M. (2000). Novel oscillators using the operational transresistance amplifier. *Microelectronics Journal, 31*(1), 39–47.
8. Cam, U., Kacar, F., Cicekoglu, O., Kuntman, H., & Kuntman, A. (2004). Novel two OTRA-based grounded immitance simulator topologies. *Analog Integrated Circuits and Signal Processing, 39*(2), 169–175.
9. Kilinc, S., Salama, K. N., & Cam, U. (2006). Realization of fully controllable negative inductance with single operational transresistance amplifier. *Circuits Systems Signal Processing, 25,* 47–57.
10. Pandey, R., Pandey, N., Paul, S. K., Singh, A., Sriram, B., & Trivedi, K. (2011). New topologies of lossless grounded inductor using OTRA. *Journal of Electrical and Computer Engineering*, 1–6, Article ID 175130.
11. Gupta, A., Senani, R., Bhaskar, D. R., & Singh, A. K. (2012). OTRA-based grounded-FDNR and grounded-inductance simulators and their applications. *Circuits, Systems, and Signal Processing, 31*(2), 489–499.
12. Gupta, A., Senani, R., Bhaskar, D. R., & Singh, A. K. (2013). New OTRA-based generalized impedance simulator. *ISRN Electronics*, 1–10, Article ID 907597.
13. Pandey, R., Pandey, N., Paul, S. K., Singh, A., Sriram, B., & Trivedi, K. (2014). Novel grounded inductance simulator using single OTRA. *International Journal of Circuit Theory and Applications, 42*(10), 1069–1079.
14. Ghosh, M., & Paul, S. K. (2014). Design of lossless grounded negative inductance simulator using single operational transresistance amplifier. *Revue Roumaine des Sciences Techniques - Serie Électrotechnique, 59,* 381–390.
15. Nagar, B. C., & Paul, S. K. (2016). Negative inductance simulator using OTRA. In *Proceedings of the IEEE international conference on microelectronics, computing and communications*, NIT Durgapur, India, 2016, 1–3, https://doi.org/10.1109/microcom2016.7522486.
16. Nagar, B. C., & Paul, S. K. (2017). Lossless grounded FDNR simulator and its applications using OTRA. *Analog Integrated Circuits and Signal Processing, 92,* 507–517.
17. Said, L. A., Radwan, A. G., Madian, A. H., & Soliman, A. M. (2015). Fractional order oscillators based on operational transresistance amplifiers. *International Journal of Electronics and Communications (AEU), 69,* 988–1003.
18. Nagar, B. C., & Paul, S. K. (2016). Voltage mode third order quadrature oscillators using OTRAs. *Analog Integrated Circuits and Signal Processing, 88,* 517–530.
19. Chang, Chun-Ming, & Swamy, M. N. S. (2013). Analytical synthesis of odd/even nth order elliptic cauer filter structures using OTRAs. *International Journal of Circuit Theory and Applications, 41,* 1248–1271.
20. Lo, Y. K., & Chien, H. C. (2006). Current-mode monostable multivibrators using OTRAs. *IEEE Transactions on Circuits and Systems-II, 53,* 1274–1278.
21. Lo, Y. K., Chien, H. C., & Chiu, H. J. (2010). Current-input OTRA Schmitt trigger with dual hysteresis modes. *International Journal of Circuit Theory and Applications, 38,* 739–746.
22. Pandey, R., Pandey, N., Chitranshi, S., & Paul, S. K. (2015). Operational transresistance amplifier based PID Controller. *Advances in Electrical and Electronic Engineering, 13,* 171–180.
23. Chien, Hung-Chun. (2016). Switch-controllable full phase operation precision half wave rectifier using a single OTRA. *Journal of Circuits, Systems, and Computers, 25,* 1–21.
24. Baker, R. J., Li, H. W., & Boyce, D. E. (1998). *CMOS circuit design, layout and simulation. Chapter 7*. New York: IEEE Press.
25. Herencsar, N., Koton, J., Jerabek, J., Vrba, K., & Cicekoglu, O. (2011). Voltage-mode all pass filters using voltage conveyor and MOSFET based electronic resistors. *Radioengineering, 20,* 10–18.
26. Yucel, F., & Yuce, E. (2015). A new voltage-mode multifunctional filter using only two voltage followers and a minimum numbers of passive elements. *Journal of Circuits, Systems and Computers, 24,* 1–16.



Bal Chand Nagar is currently working as Assistant Professor in the Department of Electronics and Communication Engineering, NIT Patna, Bihar, India. He did her B.E. (ECE) from MIT Mandsaur (RGPV, Bhopal), MP, India, M.Tech. (Microelectronics & VLSI Design) from SGSITS Indore (RGPV, Bhopal), MP, India, and Ph.D. from IIT(ISM) Dhanbad, JH, India. He has served, SVCE Indore, AITR Indore in various capacities. His research interests include Analog/Digital VLSI Circuit Design. He has published papers in International, National Journal of repute and Conferences. He is life member of IETE, ISTE, IEI and student member of IEEE for last 4 years.



Sajal K. Paul did his B.Tech., M.Tech., and Ph.D. in Radio Physics and Electronics from the Institute of Radio Physics and Electronics, University of Calcutta. He has served Webel Telecommunication Industries, Kolkata; Indira Gandhi National Open University (IGNOU), Kolkata; Advanced Training Institute for Electronics & Process Instrumentation (ATI-EPI), Hyderabad; North Eastern Regional Institute of Science & Technology(NERIST), Nirjuli and Delhi College of Engineering(DCE), Delhi in various capacities. He has served the Department of Electronics Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad as Head of the Department, Dean (Faculty) and at present Professor of the same department. His research interest includes Microelectoncic Devices, Electronic Properties of Semiconductor and Bipolar and MOS Analog Integrated Circuits. Dr. Paul has more than 115 research publications in International and National journals of repute and conferences.

Analog Integrated Circuits and Signal Processing

21st Australasian Fluid Mechanics Conference
Adelaide, Australia
10-13 December 2018

Numerical Investigation of Autoignition of Singlecomponent and Multicomponent surrogate of Diesel fuel

Anirudh Singh[1] and Amit Pal[1]

[1]Department of Mechanical Engineering
Delhi Technological University, Delhi 42, India

Abstract

This paper deals with designing of a three-dimensional computational model of an ignition quality tester and incorporating physical properties of surrogate fuel such as density, enthalpy, boiling point as a function of temperature in CONVERGE CFD to accurately and efficiently reproduce the ignition behaviour of single and multi-component surrogate of diesel fuel under different temperatures. N-Heptane is a most common diesel fuel surrogates and is also used for validating ignition quality tester. The pressure inside the chamber was fixed at 2.1MPa and the simulations were carried out at a chamber temperature of 770K, 800K, 830K and 860K. The collision, coalescence and evaporation of the fuel droplets was modelled using Patterson and Reitz-Helmholtz Rayleigh-Taylor (KH-RT) model. Skeletal chemical mechanism consisting of total 42 species and 168 reactions was used for combustion modelling of n-heptane and a reduced chemical mechanism consisting of 163 species and 887 reactions was used for determining the ignition delay for the multicomponent surrogate (n-dodecane/m-xylene) of diesel fuel to reduce the computational load. The various spray parameters like Sauter mean diameter, spray tip penetration length and vapor penetration were studied in all cases. The results show that the total ignition delay increases with the decrease in the chamber temperature. Also, it was observed that the chamber temperature has more effect on spray tip penetration over vapor penetration.

Introduction

Ignition quality of the diesel-type fuels is defined by its cetane number which can be determined either by engine-based cetane number testing methods or by non-engine-based ignitability testing methods. While the reliable measurements of these biogenic fuels have not been possible by the engine-based test methods [4], non-engine test methods like Ignition quality tester proves to be very accurate in determining the ignition quality of diesel and alternative diesel fuels with high repeatability and reproducibility.

Ignition quality tester is a bench-scale in which combustion of fuel takes place in a defined volume (0.21L) and contains a spray injection system designed for direct measurement of liquid fuel ignition delay. Ignition quality tester determines the ignition delay by measuring the time interval between the SOI (start of injection) and SOC (start of combustion) and for the measured ignition delay time, DCN (derived cetane number) is estimated using a standard correlation determined according to ASTM D6890-10a.

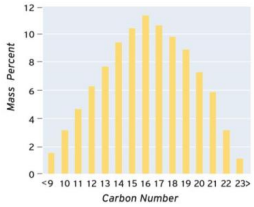


Figure 1. Carbon Number Distribution of 2-D Diesel Fuel

Diesel fuel composition

Diesel is a mixture of thousands of hydrocarbons mostly with carbon number between 10 and 22. The carbon number distribution for No. 2-D diesel fuel can be seen in Fig. 1. Most of these compounds are members of paraffin, naphthenic, or aromatic class of hydrocarbon with different chemical and physical properties [3] and thus development of chemical kinetic model for diesel is a difficult task due to its complex composition. To solve this problem, surrogate fuel with similar physical and combustion properties are develop which are used to simulate combustion and spray process of these complex fuels and hence reduce the computational time drastically.

Many researchers have tried to study the behaviour of n-heptane, n-decane, n-tetradecane and n-hexadecane under various pressure and temperature conditions as a possible surrogate for diesel fuel [7,8]. N-Heptane is most common diesel fuel surrogates and is also used for validating IQT (ignition quality tester). Numerous reduced chemical mechanisms are available in the literatures that have been validated experimentally in the shock tube experiments. Very less information is available on the how these surrogates behave under various temperature and pressure conditions.

This paper deals with a temperature sweep test conducted computationally on n-heptane and a multicomponent surrogate with a composition of 77% of n-dodecane and 23% m-xylene by volume to determine their behaviour under different temperature at a constant chamber pressure of 2.1MPa.

Methodology

A 3D model of high pressure and high temperature finite volume chamber was created in CONVERGE which can be seen in fig. 2. The pressure inside the chamber was fixed at 2.1MPa and the simulations were carried out at a chamber temperature of 770K, 800K, 830K and 860K. Thus, in total 4 simulations were carried on n-heptane and similarly another 4 simulations were carried for multicomponent fuel surrogate at an injection pressure of 20MPa and injection pulse width of 1.44ms.

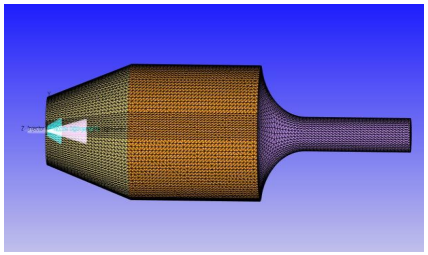


Figure 2. 3D model of ignition quality tester. Here the surface triangles should not be confused with the computational mesh.

Mesh Parameters

A three-dimensional computational model of ignition quality tester was developed on CONVERGE CFD to accurately and efficiently reproduce the ignition behaviour of single component surrogate of diesel fuel. CONVERGE uses cutcell cartesian method for mesh generation. The mesh generation process in CONVERGE is completely automatic and requires negligible human intervention for defining some basic mesh parameters like base mesh size, maximum embedding level etc.

All the simulations were carried with a base grid size of 4mm in all three directions and an adaptive mesh refinement of Level-2 based on velocity and temperature changes inside the combustion domain was employed. Adaptive mesh refinement is a very convenient technique to controls the number of elements and to provide refinement in selective areas where physical properties changes at a very fast rate. Along with the adaptive mesh refinement, a fixed embedding at a scale of 2 was also employed at the nozzle exit to resolve complex flow behaviour of spray in that region. The fixed embedding near the injector nozzle exit can be seen in the fig. 3.

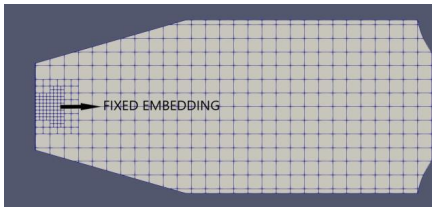


Figure 3. Fixed embedding near the injector nozzle exit

Spray Modeling and Chemical Kinetics

For liquid sprays, CONVERGE uses the Lagrangian solver to model discrete parcels and the Eulerian solver to model the continuous fluid domain. The fuel is introduced into the domain from the injector in form of parcels, that is a collection of drops. Each parcel represents a group of identical drops with same radius, velocity, temperature, etc. solves for radius, velocity, etc. The software uses sub-grid models for the spray processes that occur on length scales that are too small to be resolved to solve for radius, velocity, etc. on a per-parcel basis rather than a per-drop basis.

The effects of turbulence due to fuel spray in the constant volume chamber of the ignition quality tester is moderate compared to those in reciprocating engine so its effect was accounted using a K-epsilon turbulence model. Physical properties of the surrogates such as density, enthalpy, boiling point as a function of temperature were incorporated to mimic the droplet collision, coalescence and evaporation of the fuel droplets using Patterson and Reitz-Helmholtz Rayleigh-Taylor (KH-RT) model which represent the spray breakup process more accurately at higher pressure [6].

The drag on the fuel droplets was accounted by dynamic drag model [5]. The droplet collision and coalescence phenomenon play a vital role in obtaining accurate results in case of dense sprays. The spray and vapor penetration length were defined as distance from the nozzle tip to axial location of 95% and 1% of fuel mass fraction respectively.

The overall ignition delay of single component (n-heptane) surrogate in IQT is dominated by chemical kinetics over the spray physics and thus reduced/skeletal chemical mechanism consisting of total 42 species and 168 reactions were used for n-heptane and similarly another reduced chemical mechanism consisting of 163 species and 887 reactions was used for determining the ignition delay for the multicomponent surrogate of diesel fuel to reduce the computational time[2]. The injector modelled for the simulations is a saturation type fuel injector [1] and the variation of rate shape with the time(s) can be seen in fig. 4.

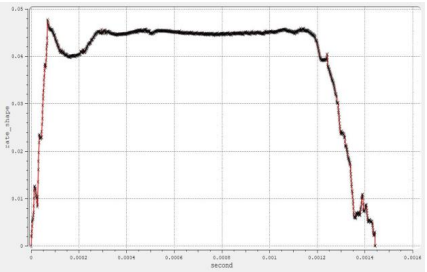


Figure 4. Injector rate shape v/s time(s)

Result and Discussion

To determine the ignition delay time, it is very crucial to define the SOI (start of injection) and SOC (start of combustion or Pressure Recovery Point). The SOI for all the simulations is set at 0ms and the SOC is defined as the time when the pressure in the chamber reaches 138KPa above the initial chamber pressure before the fuel injection [2]. The simulation was carried for 10ms with an initial time step of 1e-6s and a minimum timestep of 1e-8s. A SAGE detailed chemistry solver was used for the combustion modelling of both the surrogates. The various parameters like chamber pressure, Sauter mean diameter, spray tip penetration and vapor penetration were studied for all cases.

Chamber Pressure

The variation of chamber pressure with time for n-heptane can be seen in fig. 5. The curves clearly show an initial decrease in pressure due to the cooling effect of the liquid spray evaporation but then the chamber pressure suddenly increases because of heat release from chemical reactions. The pressure curve clearly distinguishes the physical and chemical ignition delay that can be seen by the sudden change in slope of the pressure curve at around 3ms. The total ignition delay (TID) of the fuel is the sum of physical ignition delay (PID) and chemical ignition delay (CID) [9]. The total ignition delay time for all the cases can be seen in table 1.

Sauter mean Diameter(SMD)

The variation of SMD with time for n-heptane can be seen in fig. 6. The curve shows that the SMD remains same in all the cases up to 3ms but an increase in the SMD with time can be seen at all the temperatures. A sudden increase in SMD was also observed during the time of autoignition.

Spray Tip Penetration

The spray penetration length was defined as distance from the nozzle tip to axial location of 95% of fuel mass fraction. The variation of spray tip penetration with time for n-heptane can be seen in fig. 7. The spray tip penetration is almost same for all the cases up to 1ms but after that point variation in spray tip penetration can be seen with maximum spray penetration of fuel found at chamber temperature of 770K. The spray tip penetration decreases with an increase in temperature and becomes completely zero at the time of autoignition.

Vapor Penetration Length

The vapor penetration length for all the cases can be seen in fig. 8. The vapor penetration length of the fuel is almost same at all the four temperatures and then immediately drops down to zero at the time of autoignition. The reason for fuel to show same behaviour in all cases up to 4ms is due to complete conversion of liquid phase of the fuel into vapor phase at such a high temperature.

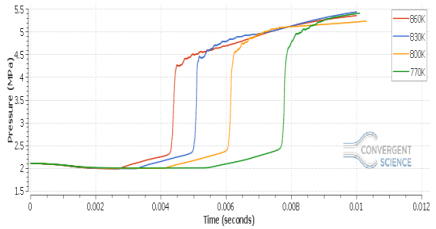


Figure 5. Variation of chamber pressure (MPa) with time(s) for n-heptane for four different chamber temperatures.

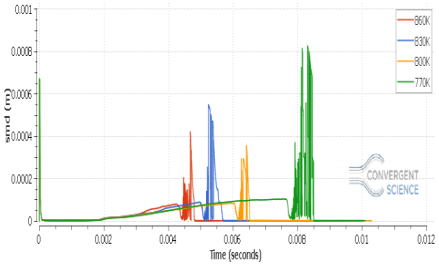


Figure 6. Variation in Sauter mean diameter(m) with time(s) for n-heptane for four different chamber temperatures.

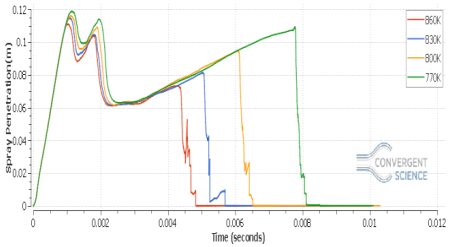


Figure 7. Variation of spray tip penetration(m) with time(s) for n-heptane for four different chamber temperatures.

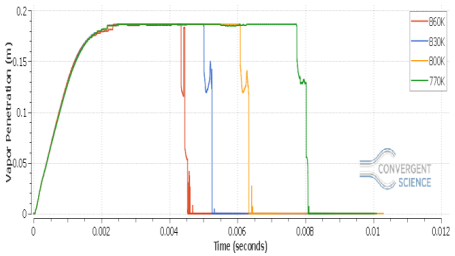


Figure 8. Variation in vapor penetration length(m) with time(s) for n-heptane for four different chamber temperatures.

S.no	Temperature(K)	TID for n-heptane (ms)	TID for multicomponent surrogate(ms)
1	770	6.58	6.32
2	800	5.28	5.10
3	830	4.34	4.12
4	860	3.71	3.47

Table 1. Total Ignition Delay for singlecomponent and multi component fuel surrogate for four different chamber temperatures.

Conclusions

A 3D model of high pressure and high temperature finite volume chamber was created to accurately and efficiently reproduce the ignition behaviour of single component and multi component surrogate of diesel fuel under different temperatures. The simulation was carried for a time interval of 10ms with an initial time step of 1e-6s and a minimum timestep of 1e-8s. The TID increases with the decrease in the chamber temperature. Also, the chamber temperature has more effect on the spray tip penetration over the vapor penetration. It was observed that the chamber pressure first decreases rapidly due to the evaporative cooling effect of liquid spray but then it increases as the chemical processes become dominant over the physical processes. During the time of autoignition, a rapid increase in chamber pressure and SMD was seen, but at the same time the spray and vapor penetration length becomes zero.

References

[1] Banish, G. Designing and Tuning High-Performance Fuel Injection Systems 2009, CarTech https://books.google.co.in/books?id=vmq7NiuaA78C

[2] G. E. Bogin Jr., A. DeFilippo, J. Y. Chen, G. Chin, Jon Luecke, M. A. Ratcliff, B. T. Zigler, and Anthony M. Dean. Numerical and Experimental Investigation of *n*-Heptane Autoignition in the Ignition Quality Tester (IQT). *Energy Fuels*, 2011, *25* (12), pp 5562–5572. October 13, 2011. DOI: 10.1021/ef201079g

[3] J. Bacha, J. Freel, A. Gibbs, L. Gibbs, G. Hemighous, K. Hoekman, J. Horn, M. Ingham, L. Jossens, D. Kohler, D. Lesnini, J. McGeehan, M. Nikanjam, E. Olsen, R. Organ, B. Scott, M. Sztenderowicz, A. Tiedemann, C. Walker, J. Lind, J. Jones, D. Scott and J. Mills, *Diesel Fuels Technical Review*, Chevron Products Company, 2007

[4] K. huber, J. Hauber "Engine bases test method for determining the ignition quality of diesel fuels" MTZ Motortechnische Zeitschrift 74(11):904-911 DOI: 10.1007/s35146-013-0266-4

[5] Liu, A., Mather, D., and Reitz, R., "Modelling the Effects of Drop Drag and Breakup on Fuel Sprays," SAE Technical Paper 930072, 1993, https://doi.org/10.4271/930072.

[6] Mubarak Ali, M., Elhagrasy, A., Sarathy, M., Chung, S. et al., "Auto-Ignition and Spray Characteristics of n-Heptane and iso-Octane Fuels in Ignition Quality Tester," SAE Technical Paper 2018-01-0299, 2018, https://doi.org/10.4271/2018-01-0299.

[7] Naser N, Yang SY, Kalghatgi G, Chung SH. Relating the octane numbers of fuels to ignition delay times measured in an ignition quality tester (IQT). Fuel 2017; 187:117–27. https://doi.org/10.1016/j.fuel.2016.09.013

[8] Xiuxiu Sun, Xingyu Liang, Hanzhengnan Yu, Yajun Wang, Zhijie Zhu "Comparison the Performance of N-heptane, N-dodecane, N-tetradecane and N-hexadecane" May 2017 Energy Procedia 105:1426-1433 DOI: 10.1016/j.egypro.2017.03.533

[9] Z. Zheng, T. Badawy, N. Henein, E. Sattler, "Investigation of physical and chemical delay periods of different fuels in the ignition quality tester," J. Eng. Gas Turb. Power 135 (2013) 061501.

Object Detection in Foggy Conditions by Fusion of Saliency Map and YOLO

[1]Sarthak Katyal
Department of Computer Engineering
Delhi Technological University
Delhi, India
sarthak.katyal1998@gmail.com

[2]Sanjay Kumar
Department of Computer Engineering
Delhi Technological University
Delhi, India
sanjay.kumar@dtu.ac.in

[3]Ronak Sakhuja
Department of Computer Engineering
Delhi Technological University
Delhi, India
ronaksakhuja@gmail.com

[4]Samarth Gupta
Department of Computer Engineering
Delhi Technological University
Delhi, India
samarthgupta1011@gmail.com

***Abstract* — Under foggy conditions, visibility decreases and causes many problems. Less visibility due to foggy conditions while driving increases the risk of road accidents. It is important to detect and recognize the nearby objects under such conditions and predict the distance of collision. There is a need to devise an object detection mechanism during foggy conditions. The paper proposes a solution to this problem by proposing a VESY(Visibility Enhancement Saliency YOLO) sensor which uses an algorithm that fuses the saliency map of the foggy image frame with the output generated from object detection algorithm YOLO (You Only Look Once). The image is sensed using image sensors in a stereo camera which are activated using a fog sensor and a depth map is generated to calculate the distance of collision. Dehaze algorithm is applied to improve the quality of the image frame whose Saliency image is generated on the basis of region covariance matrix. YOLO algorithm is also implemented on the improved quality image. The proposed fusion algorithm gives the bounding boxes of the union of the objects detected by Saliency Map and YOLO Algorithm thus proving to be a viable solution for real-time applications.**

Keywords — Saliency map, YOLO, depth map, dehaze, fusion.

I. INTRODUCTION

YOLO Architecture is a fully convolutional neural network which translates image pixels to coordinates of the generated bounding boxes and probabilities of class[7]. It is trained on full images for optimising detection. The network has a speed of 150 frames per second on Titan X GPU. Thus, YOLO can be used for streaming video in real time. YOLO has some limitations in predicting bounding boxes as every matrix element can only predict 2 boxes and can have 1 class thus limiting the number of objects nearby that can be predicted. Also, YOLO algorithm doesn't give the desired solution when working under severe conditions that affect image quality like foggy conditions. Foggy conditions make the image distorted and hence they are not detected by the algorithm. As shown in Fig. 1 (a) and (b), YOLO algorithm is unable to detect all the objects under foggy conditions. In Fig. 1 (a), One Vehicle was not detected by the general YOLO algorithm. Similarly, in Fig. 1 (b) many vehicles were left

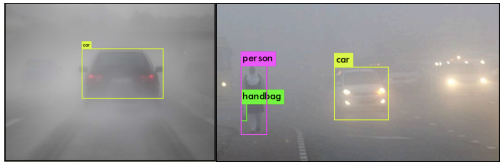


Figure 1 (a) and (b) : Failure of YOLO Detection Algorithm to distinct all objects

unidentified. This points to the fact that there is a need to enhance the existing image detection algorithm YOLO such that it can be used in foggy conditions.

For a distorted frame, like an image in Fig. 1 (a), it is important to identify each pixel's unique quality. Thus, Saliency map is generated to represent the image in a more meaningful manner. It extracts the contours from the image and this tool of image segmentation highlights the objects that are not clear in the image frame in a better manner. It can be clearly shown that the objects that were not detected and recognized by YOLO in Fig. 1 (a) and (b) have clear and distinct boundaries in their respective Saliency maps. The Saliency Map of the images have been shown in Fig. 2 (a) and (b). The intensities of white shaded regions for each pixel the saliency map points to the fact that an object exists. These intensities are incorporated in the Saliency matrix.



Figure 2 (a) and (b) : Saliency Map of the Images used in Fig 1 (a) and (b)

978-1-5386-5147-6/18/$31.00 ©2018 IEEE

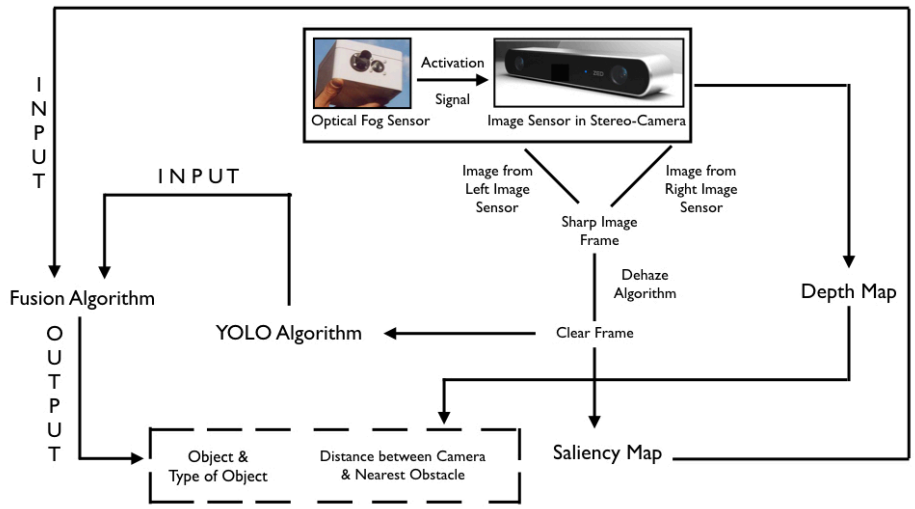


Figure 3: Prototype of VESY Sensor (Visibility Enhancement Saliency YOLO Sensor)

The results in Fig. 2 (a) and (b) thus provides additional proof to our notion of using Saliency Map for detecting objects that were not detected by YOLO algorithm. In this paper, we propose the VESY sensor which uses an algorithm to enhance the foggy image, by using the results yielded by the implementation of YOLO and the creation of saliency map for object detection and recognition during foggy conditions.

II. RELATED WORK

Reference [1] proposed a fast image defogging algorithm. We have incorporated this algorithm to obtain clear image frames. Reference [2] devised a recent method for salient object detection on images with less visibility. We have used the covariance feature matrix according to the attributes of images with fog in our proposed algorithm. The accuracy and adaptability of the method under foggy conditions is an advantage to our advanced study of fusion for better implementation of YOLO under such conditions. Reference [3] proposed a method for creation of a depth map from a stereo sensor for calculating the distance of collision. Reference [7] provides a detailed insight of You Only Look Once: Unified, Real-Time Object Detection Mechanism. The implementation fails to give desired results under adverse conditions and therefore our fusion algorithm optimises the result thus making YOLO more adaptive in nature and broader in usage to give the desired results.

III. PROPOSED WORK

A. Overview

This research yields a solution for object detection under foggy conditions on the sensed image and proposes the method to compute the distance between the object and the viewer. Fog sensor is used to activate the system and once the presence of an object is detected by our proposed algorithm, a bounding box is made around the object, and the type of the object is displayed similar to the YOLO Algorithm. Sometimes, during foggy conditions - obstacles are not clear to naked eyes, but Saliency map shows the obstacle in a distinct manner. The pre-trained model was used for the implementation of YOLO object detection algorithm. The dataset comprised of potential objects encountered as obstacles on roads especially while driving a vehicle. Sharp RGB image frames and the saliency map of the image frame were fused together. We have proposed our algorithm for this fusion process that takes into consideration the bounding boxes generated if the object is found in YOLO predicted by Saliency map and the average confidence scores for bounding boxes higher than a threshold in the saliency map in the region predicted by YOLO. Ultimately, union of all the bounding boxes is done to obtain the optimum result. The prototype of the solution has been shown in Fig. 3.

B. Optical Fog Sensor

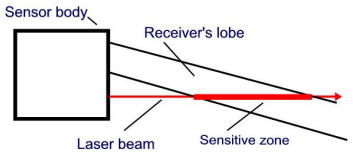


Fig 4 : Working of Optical Fog Sensor

The working of an optical fog sensor is based on backscatter technique. If the sensor gives positive output i.e. fog is sensed in the nearby environment, then the image sensor in the stereocamera is activated. Thus, the fog sensor acts as an activation signal in the overall VESY sensor. The sensor measures the amount of water particles in the air. Fig. 4 shows that a beam of laser light comes out of the sensor. There is a detector which is sensitive to incoming light in a narrow region that overlaps the transmitter beam. If there are fog particles in the overlap zone then the light will be scattered back. Light will reach the detector and send an activation signal to the image sensor. Little fog in the surrounding is enough to scatter the light and activate the image sensor. The image left and right sensors gives a clear image frame along with the depth map for the image. The image frame and the depth map are processed further before implementation of our proposed algorithm.

C. Depth Map

Using a stereo sensor, it is possible to infer the distance between two points with different depths in space having the same projection in the sensed image. Fig. 5 shows that coordinates in the target image have different positions. Based on a triangular relationship between the points, it is possible to infer the distance between the points. Fig. 5 contains equivalent triangles therefore the following result can be obtained by their equivalent equations :

$$\text{Disparity Value} = x - x' = \frac{Bf}{Z} \quad (1)$$

x' and x in equation (1) are the distance between points in image plane corresponding to the scene point 3D and their camera centre. f is the focal length of camera and distance between the two cameras has been represented by B. According to the equation (1) the depth of a coordinate in a scene has an inverse relation to the difference in distance of corresponding image coordinate and their camera centres. Using this data, we have derived the depth of all pixels in an image

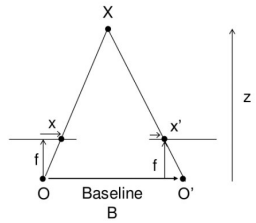


Figure 5 : Calculating Distance between object and viewer

D. Dehaze Algorithm

Dehaze algorithm was applied to each image frame obtained from the stereo sensor and the enhanced frame thus obtained was processed further to obtain the Saliency map. The algorithm is a three step process. Initially it estimates the intensity of the atmospheric light by searching in the sensed distorted image. Then the transmission map is calculated by describing a coarse map from a fine map. Finally, it will produce a clear image from the foggy image that we passed by using value of estimated intensity and transmission map. The results can be observed in Fig. 6 where the foggy image is enhanced by application of Dehaze algorithm. Initially, intensity of atmospheric light A is estimated from image I(x). We find the topmost 0.1% pixels according to their brightness in the dark channel. Pixel with highest intensity is chosen as the representative of atmospheric light. Transmission map t(x) is then estimated using using A and I(x). We first find a dark channel in regard to on a local area (Coarse Map) and then t(x) by :

$$t(x) = 1 - \text{defoggingParam} * \text{darkPixelFromCoarseMap} / \text{AtmosphericLightIntensity} \quad (2)$$

The defoggingParam in (2) is a value between 0 to 1. The higher value the lesser amount of fog would be kept for the distant objects.We can clarify the images by using the equation that follows:

$$K(x) = (I(x) - A) / \max(t(x), t(0)) + A \quad (3)$$

Equation (3) uses the following variables : K(x) is output, I(x) is input, t(x) is transmission map, A is atmospheric light and t(0) is set to a particular value to avoid dividing by zero.



Figure 6 : Result of Dehaze Algorithm to enhance Foggy Image



Figure 7 : Saliency Map for Fig. 5

E. Saliency Map

The enormous amount of information especially visual that is being sensed is so vast that various mechanisms are harnessed for choosing the most suitable part of the visual data under consideration and the rest of the part is often discarded. Reference [10] uses 2 state of the art static saliency mechanisms which are dependent on covariance based on region to process extra information, related to the depth that is available in RGB-D images. RGB-D image is an image and comprises of 4 channels. The first 3 channels of RGB-D form the general RGB image, whereas the last channel or the fourth one denotes a depth channel in alignment to the RGB component. We extend the CovSal saliency models [16] and implement them on RGB- D images. These models have been implemented on RGB image dataset which predicts saliency by accurately guessing the centre surrounding differences on the basis of 1st and 2nd order feature statistics. In our work, these models have been employed for estimating two different types of saliency maps, one from the general RGB image whereas the second one from the depth image. RGB and depth saliency maps are then combined thus giving a single saliency image for the given RGB-D as output. When compared with other state of art methods, the proposed technique of Region Covariance Matrix shows better versatility and accuracy for saliency detection of objects in foggy images. Fig. 7 shows the generated Saliency Map on one of our image frames. This result has been obtained by processing the image obtained after application of Dehaze algorithm.

F. YOLO Real Time Object Detection

The input image is split into S × S boxes by YOLO. The cell within which the centre of the object falls leads to the detection of that item. Each element in the grid estimates B number of bounding boxes and confidence values associated with every box. The confidence values shows how sure the hypotheses is that the bounding box contains that item. Confidence has been given a proper definition, as Pr(Object) ∗ IOU(Intersection Over Union). Confidence scores should be zero if no identifiable object exists in the box. Else we aim at getting the confidence value to be equivalent to the IOU between box that was predicted and the ground truth assumed. Every bounded box comprises of five predictions: x, y, w, h, along with the confidence. The mid point of the box is usually showcased by the (x,y) coordinates. (w,h) parameters are used as an alias for the image width and height. The IOU achieved between the box that was predicted and box representing ground truth gives the final confidence prediction. Each cell of

Layer (type)	Output Shape	Param #	Connected to
convolution2d_1 (Convolution2D)	(None, 16, 448, 448)	448	convolution2d_input_1[0][0]
leakyrelu_1 (LeakyReLU)	(None, 16, 448, 448)	0	convolution2d_1[0][0]
maxpooling2d_1 (MaxPooling2D)	(None, 16, 224, 224)	0	leakyrelu_1[0][0]
convolution2d_2 (Convolution2D)	(None, 32, 224, 224)	4640	maxpooling2d_1[0][0]
leakyrelu_2 (LeakyReLU)	(None, 32, 224, 224)	0	convolution2d_2[0][0]
maxpooling2d_2 (MaxPooling2D)	(None, 32, 112, 112)	0	leakyrelu_2[0][0]
convolution2d_3 (Convolution2D)	(None, 64, 112, 112)	18496	maxpooling2d_2[0][0]
leakyrelu_3 (LeakyReLU)	(None, 64, 112, 112)	0	convolution2d_3[0][0]
maxpooling2d_3 (MaxPooling2D)	(None, 64, 56, 56)	0	leakyrelu_3[0][0]
convolution2d_4 (Convolution2D)	(None, 128, 56, 56)	73856	maxpooling2d_3[0][0]
leakyrelu_4 (LeakyReLU)	(None, 128, 56, 56)	0	convolution2d_4[0][0]
maxpooling2d_4 (MaxPooling2D)	(None, 128, 28, 28)	0	leakyrelu_4[0][0]
convolution2d_5 (Convolution2D)	(None, 256, 28, 28)	295168	maxpooling2d_4[0][0]
leakyrelu_5 (LeakyReLU)	(None, 256, 28, 28)	0	convolution2d_5[0][0]
maxpooling2d_5 (MaxPooling2D)	(None, 256, 14, 14)	0	leakyrelu_5[0][0]
convolution2d_6 (Convolution2D)	(None, 512, 14, 14)	1180160	maxpooling2d_5[0][0]
leakyrelu_6 (LeakyReLU)	(None, 512, 14, 14)	0	convolution2d_6[0][0]
maxpooling2d_6 (MaxPooling2D)	(None, 512, 7, 7)	0	leakyrelu_6[0][0]
convolution2d_7 (Convolution2D)	(None, 1024, 7, 7)	4719616	maxpooling2d_6[0][0]
leakyrelu_7 (LeakyReLU)	(None, 1024, 7, 7)	0	convolution2d_7[0][0]
convolution2d_8 (Convolution2D)	(None, 1024, 7, 7)	9438208	leakyrelu_7[0][0]
leakyrelu_8 (LeakyReLU)	(None, 1024, 7, 7)	0	convolution2d_8[0][0]
convolution2d_9 (Convolution2D)	(None, 1024, 7, 7)	9438208	leakyrelu_8[0][0]
leakyrelu_9 (LeakyReLU)	(None, 1024, 7, 7)	0	convolution2d_9[0][0]
flatten_1 (Flatten)	(None, 50176)	0	leakyrelu_9[0][0]
dense_1 (Dense)	(None, 256)	12845312	flatten_1[0][0]
dense_2 (Dense)	(None, 4096)	1052672	dense_1[0][0]
leakyrelu_10 (LeakyReLU)	(None, 4096)	0	dense_2[0][0]
dense_3 (Dense)	(None, 1470)	6022590	leakyrelu_10[0][0]

Total params: 45,089,374
Trainable params: 45,089,374
Non-trainable params: 0

Figure 8 : Layers in YOLO Implementation

the grid predicts conditional probabilities of class, Pr(Class i | Object). The probabilities have been conditioned on the cell in grid which contains the desired object. One part of probabilities of class are predicted in each cell of grid as defined, nevertheless of the count of boxes. During test time multiplication of conditional class probabilities is done and confidence predictions of every box was made individually.

$$\Pr(\text{Classi}|\text{Object}) * \Pr(\text{Object}) * \text{IOU} = \Pr(\text{Classi}) * \text{IOU} \quad (4)$$

Equation (4) gives the specific confidence values for each class for every box. Probabilities of those specific classes are shown inside the box and depicts how nicely the box positions are predicted for the object under consideration. The neural network has 9 convolutional layers. After these layers, 3 fully connected layers are added to make it faster. YOLO is optimised as it predicts many bounding boxes for every cell in grid. One bounding box predictor is given the responsibility for each object at training time. The algorithm assigns one predictor the task for detecting an object. The predictor does so by finding out the prediction which has the highest current intersection over union with the assumed ground truth. Specialisation is done between the bounding box predictors because of this. Overall recall is improved as performance of every predictor is enhanced for predicting certain sizes or

aspect ratios. Fig. 8 shows the 9 convolutional layers followed by 3 fully connected layers used in our implementation. Implementation of YOLO detection algorithm on foggy images is depicted in Fig. 1. It is found that only some nearby and clear objects are detected by the object detection algorithm. Bounding boxes were not formed for the unclear objects. The other objects were clearly evident in the saliency map of the image and hence the saliency map can be used to form bounding boxes for the remaining undetected objects.

G. VESY Algorithm

The Saliency map generated is fused with results of YOLO algorithm. The following terminology has been used :
P(Yolo|Saliency) : Probability that object is found by YOLO in the region predicted by Saliency Map.
P(Saliency|Yolo) : Mean of Pixel Values of the Saliency Map in the region predicted by YOLO.

Proposed Algorithm

Variables Used : μ_s (Threshold for Saliency Map), u_y (Threshold for the confidence value of YOLO to mark bounding box), $A_{s[nxm]}$ (Auxiliary Array of the Dimensions of the Saliency Map), $S_{[nxm]}$ (Saliency Matrix), Y (YOLO produced Matrix), Y' (YOLO produced Matrix at a low threshold), R_s [lt, rt, lb, rb] (Represent the corners of square sub matrices), μ_{ys} (Threshold for finding Bounding Boxes for objects found in YOLO in the region predicted by Saliency Map), μ_{sy} (Threshold for finding Bounding Box for objects with pixel values >= μ_s in the Saliency Map in the region predicted by YOLO).

Step 1 : Create Saliency Matrix S with intensities of white regions for each pixel of the sensed image.

Step 2 : Run YOLO and generate matrix Y at threshold u_y and Y' (at a low threshold) which comprises of 5 predictions: x, y, w, h, and confidence values.

Step 3 : P(Yolo|Saliency) is computed:

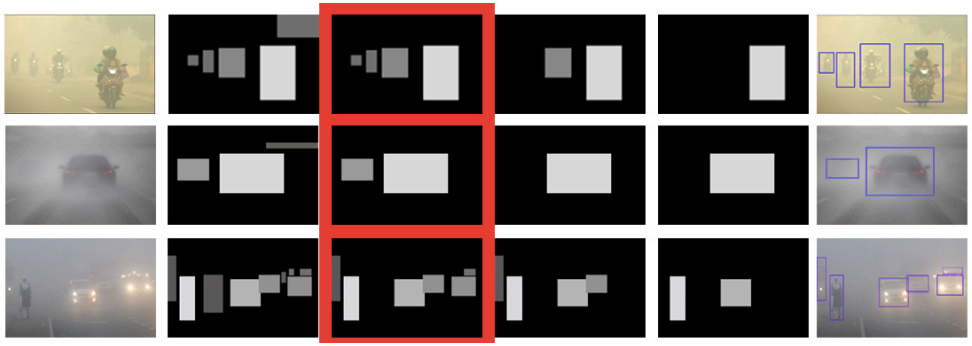
```
1   for i = 0 to n-1
2       for j = 0 to m-1
3           if Sij >= μs
4               Aij=1
5   find the max size square sub-matrices with
    Aij=1 and store their dimensions in Rs
6   for i = 0 to Rs.length - 1
7       a, b, c, d = Rs[i].lt, Rs[i].rt, Rs[i].lb, Rs[i].rb
8       sal_avg = Mean (Sij bounded by a, b, c, d)
9       y_avg  = Find Average Confidence Value in
                 Y'  for Box bounded by a, b, c, d
10      if (mean( sal_avg, y_avg) > μys)
11          mark_box  (a, b, c, d)
```

Step 4 : P(Saliency|Yolo) is computed

```
1   for i = 0 to Y.length - 1

2       a, b, c, d = Y[i].lt, Y[i].rt, Y[i].lb, Y[i].rb
3       y_avg  = Y[i].confidence
4       sal_avg = Mean (Sij bounded by a, b, c, d)
5       if (mean( sal_avg, y_avg) > μsy)
6           mark_box  (a, b, c, d)
```

Step 5 : The required output is Union of the Bounding Boxes generated by P(Yolo|Saliency) & P(Saliency|Yolo) i.e. *(P(Saliency|Yolo) U P(Yolo|Saliency))*. The above computation provides all the Bounding Box that cross the Bounding Function specified in their respective algorithms. Bounding Function in both the computations is chosen as the Threshold Values (μ_{sy} & μ_{ys}). The bounding boxes that cross the threshold our considered a part of the desired solution.



(a) (b) (c) (d) (e) (f)

Figure 9 : Results obtained after implementation of our algorithm

IV. RESULT AND DISCUSSION

The value of threshold μ_{sy} is kept fixed and results are computed by varying the value of threshold μ_{ys} for various images. The results have been computed by keeping the value of threshold μ_{ys} as 0.60, 0.70, 0.80 and 0.90. The algorithm proposed in the previous section was tested on several images and the results for three images have been showcased in Figure 9. Fig. 9 (a) shows the original image used for the implementation of the algorithm. Fig. 9 (b) shows the bounding boxes at μ_{ys} = 0.60, Fig. 9 (c) shows the bounding boxes at μ_{ys} = 0.70, Fig. 9 (d) shows the bounding boxes at μ_{ys} = 0.80, Fig. 9 (e) shows the bounding boxes at μ_{ys} = 0.90 and Fig. 9 (f) shows the final output on the input image.

It is observed that when the threshold is kept at a low value, all the objects are detected and their bounding boxes are generated even if their presence is not distinct in either YOLO object detection or the saliency map. μ_{ys} = 0.70 gives an optimum result yielding bounding boxes for all the objects in all the three images. These bounding boxes are a result of union of the bounding boxes generated by YOLO object detection and saliency map. Thus, μ_{ys} = 0.70 is a suitable value for the threshold. The result also shows that not all objects are able to cross the high threshold and thus only few distinct bounding boxes are generated in such cases. Therefore, μ_{ys} = 0.80 is not a suitable value for the threshold. Similarly, μ_{ys} = 0.90 has even a higher value of threshold and thus yields poor results. Therefore, μ_{ys} = 0.90 is not a suitable value for the threshold.

V. CONCLUSION

The union of the bounding boxes generated by YOLO algorithm and the Saliency map is achieved at the desired threshold value using the proposed algorithm. The final output detects all the objects in a foggy image frame. The proposed VESY(Visibility Enhancement Saliency YOLO) algorithm is capable of detecting additional objects that the YOLO algorithm was unable to detect. Also, distance can be calculated via the depth map and can be displayed in addition to the bounding boxes around the objects for application in Real-Time Systems.

REFERENCES

1. Zhiming Tan, Xianghui Bai, Bingrong Wang and Akihiro Higashi, "Fast Single-image Defogging" FUJITSU Sci. Tech. J., vol. 50, No. 1, January 2014.
2. Lu W., Sun X., Li C. , "A New Method of Object Saliency Detection in Foggy Images", Image and Graphics 8th International Conference, ICIG 2015, Tianjin, China, August, 2015, vol. 9217, pp.206-217.
3. Olga Krutikova, Aleksandrs Sisojevs and Mihails Kovalovs, "Creation of a Depth Map from Stereo Images of Faces for 3D Model Reconstruction", Procedia Computer Science, 2017, vol. 104, pp. 452-459.
4. P. R. Induchoodan, M. J. Josemartin and P. R. Geetharanjin, "Depth recovery from stereo images," 2014 International Conference on Contemporary Computing and Informatics (IC3I), Mysore, 2014, pp. 745-750.
5. H. Song, Z. Liu, H. Du, G. Sun, O. Le Meur and T. Ren, "Depth-Aware Salient Object Detection and Segmentation via Multiscale Discriminative Saliency Fusion and Bootstrap Learning," in IEEE Transactions on Image Processing, vol. 26, no. 9, pp. 4204-4216, Sept. 2017.
6. Guo, Jingfan & Ren, Tongwei & Bei, Jia & Zhu, Yujin. (2015), "Salient object detection in RGB-D image based on saliency fusion and propagation", The 7th International Conference, August 2015.
7. Redmon Joseph, Divvala Santosh, Girshick Ross, Farhadi Ali, "You Only Look Once: Unified, Real-Time Object Detection", eprint arXiv: 1506.02640, June, 2015.
8. Fan Guo, Jin Tang, Zixing Cai, "Fusion Strategy for Single Image Dehazing", International Journal of Digital Content Technology and its Applications(JDCTA), vol. 7, January 2013.
9. S. Ansia, A.L. Aswathy, "Single Image Haze Removal Using White Balancing and Saliency Map", Procedia Computer Science, 2015, vol. 46, pp. 12-19.
10. Erkut Erdem, "A Region Covariances-based Visual Attention Model for RGB-D Images", International Journal of Intelligent Systems and Applications in Engineering, vol. 4, pp. 128-134, December, 2016.
11. Wang ST., Zhou Z., Qu HB., Li B. (2017), "Visual Saliency Detection for RGB-D Images with Generative Model", Computer Vision – ACCV 2016. ACCV 2016, Lecture Notes in Computer Science, vol 10115. Springer, Cham.
12. R. Cong, J. Lei, C. Zhang, Q. Huang, X. Cao and C. Hou, "Saliency Detection for Stereoscopic Images Based on Depth Confidence Analysis and Multiple Cues Fusion," in IEEE Signal Processing Letters, vol. 23, no. 6, pp. 819-823, June 2016.
13. Jiawei Xu, Shigang Yue, "Fusion of Saliency Maps for Visual Attention Selection in Dynamic Scenes", International Journal of Advanced Research in Artificial Intelligence(IJARAI),2013, vol. 2.
14. L. Qu, S. He, J. Zhang, J. Tian, Y. Tang and Q. Yang, "RGBD Salient Object Detection via Deep Fusion," in IEEE Transactions on Image Processing, vol. 26, no. 5, pp. 2274-2285, May 2017.
15. Juan Du, "Understanding of Object Detection Based on CNN Family and YOLO", Journal of Physics: Conference Series, 2018, vol. 1004.
16. E. Erdem, and A. Erdem , "Visual saliency estimation by nonlinearly integrating features using region covariances", Journal of Vision, vol. 13, pp. 1-20.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 1 (2019) pp. 313-315
© Research India Publications. http://www.ripublication.com

Optimization of Material Removal Rate in Wire-EDM using Genetic Algorithms

G Daniel[1], Dr. Girishkumar[2], Dr Ali-Ul-Hasan-Rizvi[1]

[1]Department of Mechanical Engineering, Al-Falah University, Dhauj, Faridabad, India.

[2]Department of Mechanical Engineering, Delhi Technological University, Delhi, India.

Abstract

The current study focus on developing mathematical model for MRR in wire EDM using artificial neural network. The experiment is designed using L27 Taguchi orthogonal array. The process parameters selected are pulse on time, pulse off time, flushing pressure, wire tension, servo voltage and wire feed rate. A feed forward neural network with back propagation algorithm is trained to find better MRR. MATLAB program is used to develop neural network model and trained model is simulated. The network model result is compared with experimental result and it find that error is minimum. Then optimum process parameter is identified using genetic algorithm. The obtained optimum parameter is experimentally verified.

Keywords: WEDM, Artificial neural network (ANN), genetic algorithm, Taguchi technique

INTRODUCTION

WEDM is thermo-electrical unconventional machining process. The material is removed by successive spark between the work piece and wire electrode. The gap between the wire electrode and workpiece is completely filled with dielectric fluid. It will flush away the removed material from the spark gap [1]. Material removal rate is an important process characteristic need to be studied in WEDM. Selection of process parameter is an important element to study the process characteristics in EDM [2]. Somashekhar et al developed a neural network model for MRR using artificial neural network in WEDM process. The well trained model was developed using MATLAB and it found effective in estimating the MRR. The well trained model is optimized using genetic algorithm [3]. Pragya Shandilya et al have modelled the optimized the selected process parameter in WEDM process using ANN and Genetic algorithm. The selected parameters were pulse-on time, pulse-off time, wire feed rate and servo voltage, [4].

Bijaya BijetaNayak and Siba SankarMahapatra optimized the selected process parameter in EDM during taper cutting of Inconel718 material. Taguchi design experiment technique was used to run the experiment. Part thickness, taper angle, pulse duration, discharge current, wire speed and wire tension were considered as a process parameter. ANN model was developed to understand the relationship between process parameter and output responses [5]. G.Sankara Narayanan and D.Vasudevan developed a functional model relating process parameter and response variables using artificial neural network (ANN) for SKD11 tool steel material in WEDM. The trial experiments were conducted to get the experimental data to train the neural network. The program for train the model is developed using Mat lab [6]. ANN was an effective technique to predict input process parameters for optimal process performance [7]. Genetic algorithm was an effective optimization technique used for machining process [8]. In this paper ANN is used for modelling and GA is employed to optimize the MRR during WEDM.

EXPERIMENTATION

The experiment is performed on Electra Spring cut 734 four axis WEDM. The experiment is designed using Taguchi method. 27 experiments are conducted at different level of input process parameter. As six input process parameters namely pulse on, pulse off time, Current, Servo voltage wire feed and fluid pressure are chosen for the experiments. Input process parameters and their levels are shown in table 1.

Table 1: Input variable and their levels

S.No	Symbol	Input process parameter	Levels		
			I	II	III
1	A	Pulse on Time- Ton (μs)	110	115	120
2	B	Pulse off time T-off (μs)	40	45	50
3	C	Flushing Pressure (kgf/cm2)- Fp	8	10	12
4	D	Wire Tension (kgf)-Wt	550	750	950
5	E	Servo Voltage (volt)- Sv	15	20	25
6	F	Wire feed rate (m/min)- Wf	6	8	10

The range of input process parameter is selected based on the literature survey and the preliminary experiments. Few parameters are fixed as available in the machine. The fixed parameters are shown in table 2.

Table 2: Fixed Parameters

S.No	Fixed parameter	Set value
1	Wire material	Molybdenum wire of diameter of 0.25 mm
2	Peak current	230 Amps
3	Pulse in peak voltage	2
4	Servo feed setting	250

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 1 (2019) pp. 313-315
© Research India Publications. http://www.ripublication.com

The response parameter material removal rate (MRR) is calculated by weighing the workpiece before and after machining with respect to time. The observed results for MRR are shown in table 3.

Table 3: Experiments using L27 orthogonal array

Run	Ton	Toff	Fp	Wt	Sv	Wf	MRR
1	110	40	8	550	15	6	6.5826
2	110	40	8	550	20	8	6.096
3	110	40	8	550	25	10	5.7664
4	110	45	10	750	15	6	6.0711
5	110	45	10	750	20	8	5.7778
6	110	45	10	750	25	10	5.4864
7	110	50	12	950	15	6	5.6018
8	110	50	12	950	20	8	5.4668
9	110	50	12	950	25	10	4.978
10	115	40	10	950	15	6	8.3965
11	115	40	10	950	20	8	7.748
12	115	40	10	950	25	10	7.2169
13	115	45	12	550	15	6	7.029
14	115	45	12	550	20	8	6.6949
15	115	45	12	550	25	10	6.338
16	115	50	8	750	15	6	5.9778
17	115	50	8	750	20	8	5.777
18	115	50	8	750	25	10	5.4474
19	120	40	12	750	15	6	8.5164
20	120	40	12	750	20	8	8.1123
21	120	40	12	750	25	10	7.9431
22	120	45	8	950	15	6	8.1194
23	120	45	8	950	20	8	7.8864
24	120	45	8	950	25	10	7.8596
25	120	50	10	550	15	6	7.6905
26	120	50	10	550	20	8	7.1651
27	120	50	10	550	25	10	6.9987

Artificial neural network (ANN)

The back propagation training algorithms has been implemented for training the neural architectures. In this study single hidden layer was used throughout analysis. It because single hidden layer is enough for back-propagation neural network to define the input-output mapping. The number of neuron in the input layer and the output layer are fixed for this study. The input process parameters and responses of Wire - EDM process are fixed for this analysis and also the mapping is to be established between input and output parameters.

Genetic algorithm

The genetic algorithm is a probabilistic local search algorithm that iteratively transforms a Population into mathematical objects, each with an associated fitness value into a new population of offspring objects. The objective of this study is to maximize the MRR

MRR = f(Ton_Toff_ Fp_Wt_Sv_ Wf)

Subject to constraints

$110 \leq Ton \leq 120$

$40 \leq Toff \leq 50$

$8 \leq Fp \leq 12$

$550 \leq Wt \leq 950$

$15 \leq Sv \leq 25$

$6 \leq Wf \leq 10$

Where Ton, Toff, Fp, Wt, Sv and Wf are pulse on, pulse off time, Current, Servo voltage wire feed and fluid pressure respectively. Genetic algorithm parameter and their value is shown in table 5. The limitation range of the input process parameters are assigned based on experimental setup. f represents relationship between material removal rate and six input process parameters. For every iteration of Genetic algorithms, material removal rate values are predicted by artificial neural network models for a specified population size. The Genetic algorithm randomly generated a real valued population of possible process parameter chromosomes. A new population for the next generation is created based on old population using different operator functions. MATLAB is used for the optimization. Optimized parameter value is shown in table 6.

Table 4: Predicted Value

S.No	MRR-Actual	ANN MRR-Predicted	S.No	MRR-Actual	ANN MRR-Predicted
1	6.5826	6.20119	15	6.338	6.76547
2	6.096	6.10152	16	5.9778	6.21164
3	5.7664	5.77725	17	5.777	5.71987
4	6.0711	6.14578	18	5.4474	5.42603
5	5.7778	5.82429	19	8.5164	8.39569
6	5.4864	5.62399	20	8.1123	8.33866
7	5.6018	5.97842	21	7.9431	8.15525
8	5.4668	5.70123	22	8.1194	8.14779
9	4.978	5.08302	23	7.8864	7.92659
10	8.3965	8.30814	24	7.8596	7.90579
11	7.748	8.03780	25	7.6905	7.83080
12	7.2169	7.58424	26	7.1651	7.24643
13	7.029	6.94590	27	6.9987	6.66245
14	6.6949	7.03865			

Table 5: Genetic Algorithm parameter and their values

number of generations	500
population size	20
crossover rate	0.8
crossover mechanism	Two point
mutation rate	0.01

Table 6: Optimum process parameter

Process parameter						MRR		Error
Ton	**Toff**	**Fp**	**Wt**	**Sv**	**Wf**	predicted	Experimental	
115	40	10	750	15	8	7.728	7.964	3.05%

DISCUSSION

ANN is trained with different nodes in the hidden layer using experimental data. Model accuracy is improved of low value of RSME model. In this study two layer and six hidden neurons per layer is selected. The predicted value of MRR is compared with the experimental model and shown in table 4. To predict MRR for nonlinear responses tansig transfer function introduced in the neural network.

The selection of important input process parameter is essential in manufacturing process because these factors determine performance of WEDM process. The result shows that artificial neural network model would be successfully applied in WEDM process to predict the machining response.

ANN- GA is combined to predict the optimum input process parameter for the WEDM process. The obtained results are shown in Table 5. The obtained results are verified with experimental results. The results are valid for the range of parameters considered in this analysis.

CONCLUSIONS

In this research work MRR was predicted by feed-forward neural network with back propagation algorithm for the given range of input process parameter. The following conclusion were made from the study,

1. The proposed ANN algorithm useful for reduction in production time and set-up time, along with the reduction in cost in WEDM processes.
2. The developed artificial neural network model results are close agreement with experimental values.
3. Optimum process parameter was derived from gentic algorithm process. The obtained optimum parameters are 115Ton of 114, Toff of 40, Fp of 10, Wt of 750, Sv of 15 and Wf of 8.

REFERENCE

[1] Ho, K.H., Newman, S.T., Rahimifard,S., & Allen, R. (2004). State of art in wire electric discharge machining (WEDM). *International Journal of Machine Tools and Manufacture*, 44, 1247-1259.

[2] Yousef, B.F., Knopf, G.K., Bordatchev, E.V., & Nikumb, S.K. (2003). Neural network modeling and analysis of the material removal process during laser machining. *International Journal of Advanced Manufacturing Technology*, 22. 41-53.

[3] Somashekhar, K.P., Ramachandran, N., & Jose Mathew.(2010). Optimization of Material Removal Rate in Micro-EDM Using Artificial Neural Network and Genetic Algorithms. *Materials and Manufacturing Processes*, 25, 467-475.

[4] Pragya Shandilya & Abhishek Tiwari. (2014). Artificial Neural Network Modeling and Optimization using Genetic Algorithm of Machining Process. *Journal of Automation and Control Engineering*, 2(4), 348-352.

[5] Bijaya BijetaNayak & Siba SankarMahapatra. (2016). Optimization of WEDM process parameters using deep cryo-treated Inconel 718 as work material. *Engineering Science and Technology, an International Journal*, 19, 161-170.

[6] Sankara Narayanan,G., & Vasudevan , D. (2014). Algorithm for Modeling Wire Cut Electrical Discharge Machine Parameters using Artificial Neural Network. *International Journal of Engineering and Technology*, 6, 164-170.

[7] Yusup, N. (2012). Evolutionary techniques in optimizing machining parameters: Review and recent applications. Expert Systems with Applications, 39, 9909–9927.

[8] N. K. Jain, V. K. Jain, and K. Deb, "Optimization of process parameters of mechanical type advanced machining process using genetic algorithm," International Journal of Machine Tools and Manufacture, vol. 47, pp. 900-919, 2007.

[9] Hargovind Soni., Narendranath., & Ramesh , R.M. (2018). ANN and RSM methods for predicting Material Removal Rate and Surface Roughnessduring WEDM of Ti50 Ni40Co10(shape memory alloy). *AMSE Journals IJETA Publication*, 54, 435-443.

[10] Shivade, A.S., Shinde, V.D. (2014).Multi-objective optimization in WEDM of D3 tool steel using integrated approach of Taguchi method & Grey relational analysis. *International Journal of Industrial Engineering, 10*, 149-162.

[11] Singh, J., Kalra, P., & Walia, R.S. (2014). Multi-response optimization of manual material handling tasks through utility concept. Bonfring. *International Journal of Industrial Engineering and Management Science, 4*, 101-107.

[12] Sreenivasa, R.M., Venkaiah, N. (2015). Parametric optimization in machining of Nimonic-263 alloy using RSM and particle swarm optimization. *Procedia Material Science Elsevier, 10*, 70-79.

Optimum Preventive Maintenance Policy for a Mechanical System Using Semi-markov Method and Golden Section Technique

G. Kumar[1], J. P. Varghese[2]
[1]Department of Mechanical Engineering, Delhi Technological University, Delhi, India
[2]eHealth Inc., Sanfrancisco, California, USA
(girishkumar@dce.ac.in; joelprince25@gmail.com)

Abstract - **This paper presents a method to determine an opportune time interval to carry out preventive maintenance for a two component series system to maximize system availability. Owing to the fact that the failure and repair distributions of most of the systems follow non-exponential distributions such as Weibull, Log-normal, Gamma, etc. and preventive maintenance is carried out at fixed intervals of time, the Markov method becomes inapplicable. The Semi-Markov approach which can handle non-exponential distributions provides a realistic analysis of such systems. The Semi Markov process model is solved to evaluate steady state system availability using the two-stage analytical approach. The resultant system availability result is validated using the Monte Carlo Simulation. The preventive maintenance interval time for optimal system availability is then determined using the Golden Section Search technique.**

Keywords – **Semi-Markov, availability, Monte Carlo simulation, preventive maintenance**

I. INTRODUCTION

Preventive maintenance is defined as the maintenance activity undertaken regularly at preselected intervals while the system is satisfactorily operating to bring the system back to the fully functional state [1]. The time interval of performing preventive maintenance plays a crucial role on the availability of the industrial system. It can be understood that too little maintenance may have very costly consequences, but on the other hand, it may not be economical to perform it too frequently [2]. A large preventive maintenance interval provides the opportunity for random failures to occur which decrease system availability, while a short preventive maintenance interval results in unnecessary stopping of the industrial system and can also result in maintenance induced failures [3], further decreasing system availability. The previously mentioned two extremes suggest that there must exist a preventive maintenance time interval for which the industrial system has maximum availability.

The Semi-Markov process (SMP) is applied for an analysis of a single component with a preventive maintenance policy [4]. Specific cases of components having failure distributions such as Hypo-exponential and Weibull were studied in detail to determine the optimal time for preventive maintenance but repair distributions in each case was assumed exponential. A Markov decision process is used to find the preventive maintenance time for maximum availability of a single component with deterioration [5]. They considered random failures and preventive maintenance intervals to be exponentially distributed. The same model was later solved using a Semi-Markov process by incorporating the mean sojourn time of each state [6]. The exponential distribution assumption was not relaxed in their analysis. SMP is also applied to evaluate optimal condition monitoring interval [7].

The existing literature uses the Markov and Semi-Markov processes to analyze a preventive maintenance policy. However, these are limited to exponential distributions to simplify the evaluation process. Furthermore, no method to validate results has been cited in the existing literature.

This paper overcomes these limitations by determining the availability of a two component series system using the Semi-Markov method. The results are validated using the Monte Carlo Simulation and the preventive maintenance time for optimal availability is determined using the Golden Section Search technique.

The remaining paper is organized as follows. Section II describes the system model. The methodology used is detailed in Section III. Section IV illustrates the methodology with an example. Finally, Section V concludes the work.

II. SYSTEM MODEL

The system studied in this research is a two component series system. Each component can represent any system such as a pump, turbine, transformer, etc. To carry out preventive maintenance on a component, it must be stopped. This stoppage leads to the stoppage of the entire system as the components are in series. To prevent unnecessary stoppage of each component individually, it becomes advantageous to carry out preventive maintenance for both components simultaneously when the system is stopped. It is assumed that the time of failure of each component follows a Weibull distribution and the time to repair follows a Lognormal distribution (Refer Table I). However, based on system characteristics the failure and repairs can be modelled by various other distributions. The components can also individually fail before the preventive time interval and the suitable corrective repair action is undertaken. The system can exist in four states (Refer Fig. 1).

State 1: This is the state of the system in which both components are fully functional and the system performs its required task.

978-1-5386-6786-6/18/$31.00 ©2018 IEEE

TABLE I
LIST OF DISTRIBUTIONS FOR THE SYSTEM MODEL

Cdf	Distribution
F_{12}	Weibull (β_{12}, θ_{12})
F_{13}	Weibull (β_{13}, θ_{13})
F_{14}	Deterministic or Fixed Interval(T_1)
F_{21}	Log-Normal (μ_{21}, σ_{21})
F_{31}	Log-Normal (μ_{31}, σ_{31})
F_{41}	Log-Normal (μ_{41}, σ_{41})

State 2: This is the state of the system when the first component has failed. It is an "Unavailable" state of the system as the failure of a component stops the system.

State 3: This state is similar to State 2 but occurs when the second component fails.

State 4: This state occurs at fixed intervals of time when the system is stopped for preventive maintenance on both the components.

III. METHODOLOGY

In the following section, the methodologies of the various approaches used in the analysis of the system model are elucidated.

3.1 *Semi-Markov method*

The consideration of non-exponential distributions such as the Weibull, Log-Normal and Deterministic makes the Markov approach inapplicable. The Semi-Markov process overcomes this limitation as this is capable of handling non-exponential distributions. The two-step analytical approach [8] has been used in this article to determine the steady state availability of the system model described in Section II is briefly detailed next.

Step 1: The steady state probabilities of the Embedded Markov Chain (EMC) of the system model are determined by evaluating the one-step transition probability matrix of the EMC.

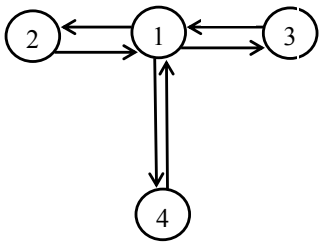


Fig. 1. System model

Step 2: The mean sojourn time spent in each state of the system is determined. The mean sojourn time of state gives an indication of the duration of stay of the system in a particular state.

The steady state probabilities of each of the system states are determined by using the values of probabilities determined in Step 1 and the mean sojourn times determined in Step 2.

3.1.1 *Step 1 –* Steady *state probabilities of the embedded Markov chain*

The first step involves the determination of the kernel matrix K(t) of the SMP model. The rows and columns of the matrix are equivalent to the number of system states. Let $k_{ij}(t)$, be an element of the kernel matrix at the i^{th} row and j^{th} column. $k_{ij}(t)$ represents the transition probability from state i to the next reachable state j within a time 't'. If a system has four states, the kernel matrix has 16 elements. Refer to "(1)."

$$K(t) = \begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} \\ k_{21} & k_{22} & k_{23} & k_{24} \\ k_{31} & k_{32} & k_{33} & k_{34} \\ k_{41} & k_{42} & k_{43} & k_{44} \end{bmatrix} \quad (1)$$

If $F_{ij}(t)$ is the corresponding CDF of the transition from state i to state j and $\bar{F}_{ij}(t)$ is the complement of the CDF, then each element of the above kernel matrix is defined as [7]:

$$k_{ij}(t) = \begin{cases} 0 & \text{When there is no possible transition from State i within the transition time t.} \\ F_{ij}(t) & \text{When there is a single possible transition from State i to State j within the transition time t.} \\ \int_0^t \bar{F}_{ik}\bar{F}_{im}\,dF_{ij}(x) & \text{When there are multiple possible transitions from State I such as State j, State k and State m within transition time t.} \end{cases}$$

The one-step transition probability matrix, Z=K(∞) is obtained using the kernel matrix by setting the steady state condition of time tending to infinity. The elements of matrix Z are represented by the term p_{ij} corresponding to each k_{ij}. To help in the evaluation of the elements of matrix Z, there is an additional condition that sum of all the elements of each row of the matrix is one [8]. Refer to "(2)."

$$\sum_{j=1}^{N} p_{ij} = 1 \quad (2)$$

Let $v=[v_1, v_2, ..., v_N]$, be the steady-state probability vector of the Embedded Markov chain. Then a system of linear equations was developed to evaluate the steady state probabilities of the Embedded Markov chain [8]. Refer to "(3)." There is also an additional condition that the sum of the probabilities of all possible states must be one. Refer to "(4)."

$$v = vZ \quad (3)$$

$$\sum v_i = 1 \quad (4)$$

Solving the set of linear equations, the steady state probabilities of the EMC can be evaluated.

3.1.2 *Step 2 – Mean sojourn time and steady state probability of the system model*

The second step involves determining the mean sojourn time for each state of the system. Let τ_i represent the mean sojourn time of state i. The mean sojourn time for each state is defined as [7]:

$$\tau_i = \begin{cases} \infty & \text{When there is no possible transition from State i within the transition time t.} \\ \int_0^\infty \bar{F}_{ij}(t)\,dt & \text{When there is a single possible transition from State i to State j within the transition time t.} \\ \int_0^\infty \bar{F}_{ij}\bar{F}_{ik}\bar{F}_{im}(t)\,dt & \text{When there are multiple possible transitions from State I such as State j, State k and State m within transition time t.} \end{cases}$$

Let P_i represent the steady state probability of state i. By using the values of the steady state probability of the EMC determined from Step 1 and the mean sojourn times of each state, the steady state probability of each state of the system can be evaluated [7]. Refer to "(5)."

$$P_i = \frac{v_i \tau_i}{\sum v_i \tau_i} \quad (5)$$

If the system consists of n "Available" states, then the steady state availability (A_s) can be determined. Refer to "(6)."

$$A_s = \sum_{i \in n} P_i \quad (6)$$

IV. ILLUSTRATION

A system of two pumps in series is chosen to illustrate the proposed methodology for determining the opportune preventive maintenance interval for maximum system availability. The parameter values for the failure and repair distributions of the pumps are selected in consultation with industry experts as well as literature [9]. Refer, Table II. The preliminary value of the preventive maintenance interval T_1 is assumed randomly for the purpose of validation of the model. After validation, the preventive maintenance interval for optimal availability can be determined using the Golden Section search technique.

Using the two-stage analytical approach described in Section II, the steady state probability of each state is determined.

The expressions of values of non-zero elements of kernel matrix, K(t) and one-step transition probability matrix, Z are derived for the system model and their values are evaluated. These are shown in Table III and Table IV respectively and included in Appendix.

Further, the values of stead-state probability of the states of EMC, values of mean sojourn time and steady state probability of system states are evaluated following the procedure detailed in Section III. These values are listed in Table V, Table VI and Table V. These results are placed in Appendix. The steady state availability of the system is given as the sum of the probabilities of the "Available" states. For the proposed system model, the only "Available" State is State 1. Hence, the evaluated value of steady-state availability is A_s = 0.992758.

To validate the result, a Monte Carlo Simulation code is formulated in MATLAB. A large mission time is selected to simulate steady state and the code is executed for 10000 trials to evaluate steady state system availability. The steady state availability based on Monte Carlo Simulation is calculated as As = 0.992809. The simulation result is shown in Fig. 2. Since there is only a paltry percentage error of 0.005%, the proposed Semi Markov model has been validated by the Monte Carlo simulation.

Using certain preliminary analysis, preventive maintenance interval for maximizing availability is estimated to be in between T_1=4500 hours to 5500 hours. Choosing the initial end points as 4500 hours and 5500 hours and applying the Golden Search Technique, the preventive maintenance interval time for optimal

TABLE II
PARAMETER VALUES FOR THE PUMP

CDF	CDF Expression	Parameter Values
F_{12}	$1-e^{-(\frac{t}{\theta_{12}})^{\beta_{12}}}$	$\beta_{12}=1.1$ $\theta_{12}=15000$
F_{13}	$1-e^{-(\frac{t}{\theta_{23}})^{\beta_{23}}}$	$\beta_{13}=1.5$ $\theta_{13}=11000$
F_{14}	$U(t-T_1)=\begin{cases}0 & t<T_1\\ 1 & t\geq T_1\end{cases}$	240 (hours)
F_{21}	$\frac{1}{2}+\frac{1}{2}erf(\frac{\ln(t)-\mu_{21}}{\sqrt{2}\sigma_{21}})$	$\mu_{21}=2.08$ $\sigma_{21}=0.59$
F_{31}	$\frac{1}{2}+\frac{1}{2}erf(\frac{\ln(t)-\mu_{31}}{\sqrt{2}\sigma_{31}})$	$\mu_{31}=2.3$ $\sigma_{31}=0.64$
F_{41}	$\frac{1}{2}+\frac{1}{2}erf(\frac{\ln(t)-\mu_{31}}{\sqrt{2}\sigma_{31}})$	$\mu_{41}=0.4$ $\sigma_{41}=0.41$

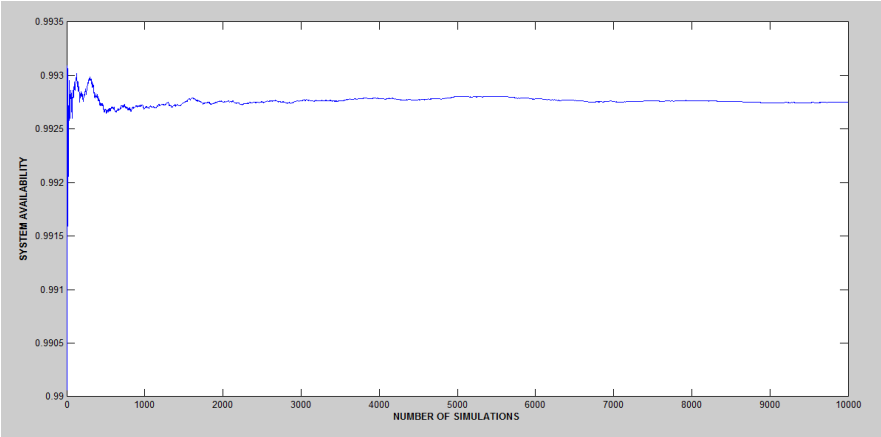


Fig. 2: Result of Monte-Carlo Simulation

availability as is obtained as T_1= 4978.6 hours with maximum system availability as $A_{maximum}$ = 0.998509.

V. CONCLUSION

In this work, a Semi-Markov process is used to determine system availability for a two component series system undergoing a preventive maintenance policy. The system model is developed by considering a Weibull failure distribution and a Log-normal repair distribution. Irrespective of the failure and repair distributions of the system, the proposed methodology can be used to determine the optimal system availability. The system undergoes preventive maintenance at fixed intervals of time, which is modelled by a deterministic distribution. The analytical result of the Semi Markov process model is validated by the Monte Carlo simulation. The preventive maintenance time interval which maximizes system availability is then evaluated using the Golden Section search technique. For components in the series configurations, performing preventive maintenance individually on separate components at different time intervals is an inexpedient tactic. It results in excessive system stoppages. It is advantageous to perform preventive maintenance on both components simultaneously while stopping the system. The suggested approach provides a powerful tool for maintenance engineers to analyze such systems and develop a suitable preventive maintenance policy to maximize the operating time of systems.

REFERENCES

[1] S. H. Sim and J. Endrenyi, "Optimal preventive maintenance with repair," *IEEE Trans. on Reliability,* vol. 37, no.1, pp. 92-96, 1988.

[2] J. Endrenyi, G. J. Anders and A.M. Leite da Silva, "Probabilistic evaluation of the effect of maintenance on reliability – an application," *IEEE Trans. on Power Systems*, vol.13, no. 2, pp. 576-583, 1998.

[3] C. E. Ebeling, *An introduction to reliability and maintainability engineering,* New Delhi, India: *Tata McGraw Hill*, 2000.

[4] D. Chen, and K. S. Trivedi, "Analysis of periodic preventive maintenance with general system failure distribution," *In Proc. 2001 Pacific Rim International Symp. Dependable Computing*, *IEEE*, Seoul, Korea, 103-107, 2001.

[5] G. K. Chan, and S. Asgarpoor, "Optimum maintenance policy with markov processes," *Electric Power Systems Res.,* vol. 76, pp. 452-456, 2006.

[6] C. L. Tomasevicz, and S. Asgarpoor, "Preventive maintenance using continuous-time semi-markov processes," In *Proc. North American Power Symp, IEEE*, Carbondale, IL, 3-8, 2006.

[7] G. Kumar, V. Jain and O. P. Gandhi, "Availability analysis of mechanical systems with condition-based maintenance using Semi-Markov and evaluation of optimal condition monitoring interval," *J. of Industrial Engineering International,* vol. !4, No. 1, pp.119-131,2018

[8] V. G. Kulkarni, *Modeling and analysis of stochastic systems,* London, UK. Chapman & Hall, (1995).

[9] A. Dennis, "Application of Monte Carlo simulation system reliability analysis," *20th International Pump Users Symp.,* Texas, USA, 2003.

Appendix

k_{ij}	$k_{ij}(t)$	$k_{ij}(t)$ with specific distribution
k_{12}	$\int_0^t \bar{F}_{14}\bar{F}_{13}dF_{12}$	$\frac{\beta_{12}}{\theta_{12}^{\beta_{12}}}\int_0^t U(x-T_1)x^{(\beta_{12}-1)}e^{-[(\frac{x}{\theta_{12}})^{\beta_{12}}+(\frac{x}{\theta_{13}})^{\beta_{13}}]}dx$
k_{13}	$\int_0^t \bar{F}_{12}\bar{F}_{14}dF_{13}$	$\frac{\beta_{13}}{\theta_{13}^{\beta_{13}}}\int_0^t U(x-T_1)x^{(\beta_{13}-1)}e^{-[(\frac{x}{\theta_{13}})^{\beta_{13}}+(\frac{x}{\theta_{12}})^{\beta_{12}}]}dx$
k_{14}	$\int_0^t \bar{F}_{12}\bar{F}_{13}dF_{14}$	$\int_0^t e^{-[(\frac{x}{\theta_{12}})^{\beta_{12}}+(\frac{x}{\theta_{13}})^{\beta_{13}}]}d(U(x-T_1))$
k_{21}	$F_{21}(t)$	$\frac{1}{2}+\frac{1}{2}erf(\frac{\ln(x)-\mu_{21}}{\sqrt{2}\sigma_{21}})$
k_{31}	$F_{31}(t)$	$\frac{1}{2}+\frac{1}{2}erf(\frac{\ln(x)-\mu_{31}}{\sqrt{2}\sigma_{31}})$
k_{41}	$F_{41}(t)$	$\frac{1}{2}+\frac{1}{2}erf(\frac{\ln(x)-\mu_{41}}{\sqrt{2}\sigma_{41}})$

TABLE III:
EXPRESSIONS OF NON- ZERO ELEMENTS OF MATRIX K(T)

TABLE IV
EXPRESSIONS AND VALUES OF THE NON-ZERO ELEMENTS OF MATRIX Z.

p_{ij}	Expressions of Elements of matrix Z, $p_{ij}=k_{ij}(\infty)$	p_{ij} values
p_{12}	$\frac{\beta_{12}}{\theta_{12}^{\beta_{12}}}\int_0^{T_1} x^{(\beta_{12}-1)}e^{-[(\frac{x}{\theta_{12}})^{\beta_{12}}+(\frac{x}{\theta_{13}})^{\beta_{13}}]}dx$	0.010511
p_{13}	$\frac{\beta_{13}}{\theta_{13}^{\beta_{13}}}\int_0^{T_1} x^{(\beta_{13}-1)}e^{-[(\frac{x}{\theta_{13}})^{\beta_{13}}+(\frac{x}{\theta_{12}})^{\beta_{12}}]}dx$	0.003198
p_{14}	$1-[k_{12}(\infty)+k_{13}(\infty)]$	0.986291
p_{21}	1	1
p_{31}	1	1
p_{41}	1	1

TABLE V
STEADY-STATE PROBABILITY OF THE STATES OF EMC/SMP

ν_i	Steady State Probabilities of the EMC	P_i	Steady State Probability of SMP
ν_1	0.500000	P_1	0.992758
ν_2	0.005256	P_2	0.000417
ν_3	0.001599	P_3	0.000163
ν_4	0.493145	P_4	0.006662

TABLE VI
EXPRESSIONS AND VALUES OF MEAN SOJOURN TIME

τ_i	Mean Sojourn Time Definitions	Mean Sojourn Time Expressions	Mean Sojourn Time Values
τ_1	$\int_0^\infty \bar{F}_{12}\bar{F}_{13}\bar{F}_{14}dx$	$\int_0^{T1} e^{-[(\frac{x}{\theta_{12}})^{\beta_{12}}+(\frac{x}{\theta_{13}})^{\beta_{13}}]}dx$	238.488104
τ_2	$\int_0^\infty \bar{F}_{21}dx$	$e^{\mu_{21}+\frac{1}{2}\sigma_{21}^2}$	9.526239
τ_3	$\int_0^\infty \bar{F}_{31}dx$	$e^{\mu_{31}+\frac{1}{2}\sigma_{31}^2}$	12.241110
τ_4	$\int_0^\infty \bar{F}_{41}dx$	$e^{\mu_{41}+\frac{1}{2}\sigma_{41}^2}$	1.622633

Defence Science Journal, Vol. 69, No. 1, January 2019, pp. 37-45, DOI : 10.14429/dsj.69.12879
© 2019, DESIDOC

Particle Swarm Optimisation of Hole Quality Characteristics in Laser Trepan Drilling of Inconel-718

Kedari Lal Dhaker[#] and Arun Kumar Pandey[*]

[#]*Mechanical Engineering Department, Jaypee University of Engineering and Technology, Guna - 473 226, India*
[*]*Department of Mechanical Engineering, Bundelkhand Institute of Engineering and Technology, Jhansi - 284 128, India*
[*]*E-mail: arunpandey76@gmail.com*

ABSTRACT

Inconel-718 is a nickel based super alloy and is extensively in use for working at very high temperature (upto 2000 °C) such as aero engine gas path equipment, nuclear equipment etc. Drilling micro size hole in such material with laser beam has been a proven choice and laser drilling process produces geometrically and dimensionally improved hole. Hole geometrical features can be improved further if laser drilling system operated at optimal input parameter setting. This paper experimentally investigates the behavior of hole geometrical features hole circularity and hole taper in laser trepan drilling of Inconel -718 sheet. Optimal value of laser input parameters for improved hole circularity and reduced hole taper have been suggested with the help of computational intelligence technique particle swarm optimisation. The effect of each laser input parameter on hole quality characteristics are also discussed and demonstrated graphically. Finally the experimental validation of the predicted results has been carried out.

Keywords: Laser trapan drilling; Inconel-718; Circularity; Hole taper; Particle swarm optimisation; Regression model

1. INTRODUCTION

Conventional drilling process like punching and twisting is not able to produces quality and precise holes in high tough and better mechanical properties materials such as Inconel 718. The high toughness and better mechanical properties of materials create high temperature zone around the drilling area and causes the thermal stresses and brittleness in worksheet and cutting tool. Therefore, the conventional drilling process is not suitable for the machining of better mechanical properties material which are being in used for technologically advanced applications like aerospace, nuclear equipment manufacturing, and medical equipment, etc[1]. Laser beam drilling, has been proven to be an efficient drilling process that can compete in terms of cost, quality and easy automation for large scale manufacturing or drilling of material such as nickel based superalloy.

Inconel-718 is a nickel based superalloy and one the most difficult to conventionally drilled alloy, in order to satisfy quality and close dimensions requirement[2]. Inconel-718 is extensively in use for working at very high temperature (upto 2000 °C) such as aero engines and gas path equipment[3]. The Generating thousands of close micro size holes in such equipment and material with laser beam has been a proven choice[4].

Two type of laser drilling process available at commercial level, laser trepan drilling (LTD) and laser percussion drilling (LPD). Laser trepan drilling, is the cutting along the circumference of the hole to be drilled, and laser percussion drilling, is the directly punching of laser beam on the material without any relative motion between work and laser beam[5]. Geometrical quality characteristics such as hole circularity and hole taper of laser trepan drilled hole are better than laser percussion drilled hole[6]. Laser trepan drilling also exhibit some geometrical and metallurgical defects such as poor circularity, tapered hole, barreling, metallurgical defects, recast layer etc[7]. Therefore improving these defects in laser trepanning drilling and getting improved geometry of hole is the major issues for manufacturer. Circularity and hole taper are the severe geometrical defects if components are being used in advanced applications such as gas turbine components, marine components, and automobile parts, etc.

The researchers have studied the laser machining and drilling processes in different ways. Pandey and Dubey[8] improved the laser cut geometry of duralumin sheet by applying the hybrid approach of robust parameter design methodology and Fuzzy logic theory. Similarly Norkey and Dubey[9] have developed mathematical model and did optimisation by genetic algorithm (GA) of heat affected zone (HAZ) in pulse-Nd:YAG laser cutting of duralumin sheet. Chien and Huo[10] have achieved the minimum recast layer thickness in laser trepan drilling by Taguchi based analysis on material Inconel-718. Goyal and Dubey[11] have improved the hole circularity and hole taper in laser trepan drilling of titanium alloy sheet by applying the regression analysis and genetic algorithm. Marimuthu[12], *et al.* have studied the Quasi-CW-fibre laser trepanning drilling of

Received : 14 May 2018, Revised : 25 October 2018
Accepted : 12 December 2018, Online published : 10 January 2019

nickel based alloy and controlled the recast layer, oxide layer, hole surface roughness and suggested the reduced recast layer thickness to 30 µm.

Tamrin[13], *et al.* had achieved lower HAZ in Laser percussion drilling of three different thermoplastics by using grey relational analysis. Hajdarevic and Bijelonja[14] had found the temperature distribution by finite element method (FEM) and experimentally validate the findings. Rajesh[15], *et al.* have developed RSM models for circularity and taper in laser drilling of austenitic stainless steel. Chaudhary[16], *et al.* have developed artificial computational based software tool for simultaneously optimisation of laser cutting quality characteristics and validate the tool with previously published results. Dhaker[17], *et al.* have investigated the laser trepan drilling on Inconel-718 and found the better drilled hole diameter by genetic algorithm.

For manufacturing process modeling and optimisation, many soft computing techniques are being in use such as neural networks (NN), fuzzy systems (FS), perception's, evolutionary computation (EC) which include Evolutionary algorithms, genetic algorithms (GA), ant colony optimisation - ACO), Hybrid (neuro – fuzzy, and Swarm intelligence (Particle swarm optimisation – PSO. Particle swarm optimisation (PSO) algorithm is relatively simple and competitive in term of performance[18]. For single objective optimisation this algorithm has been favored when dealing with many real-world optimisation problems[19]. PSO is easy in implementation and can be applied in continues as well discrete variable problem, PSO is emerging as better suitable for engineering system in term of its efficiency and more effectiveness[20].

After studying the available research literature on laser drilling, it has been concluded that conventional drilling is being replacing by laser drilling processes for getting improved hole quality. Geometry of laser drilled holes can be improved if laser process parameters are being set at their optimal levels. Laser input parameters such as assist gas pressure, laser current, stand-off distance, trepanning speed and pulse frequency, greatly affect the hole geometry. The authors find very little research available in laser trepan drilling and could not find any research paper on PSO based optimisation of laser trepan drilling process.

Research presented here is an effort to further add to an existing knowledge in laser trepan drilling process and suggest optimal values of laser input parameters for improved hole circularity and reduced hole taper in laser trepan drilling of extensity used Inconel-718 material. The experiments have been performed as per one parameter at time (OPAT) variation. A second order regression model has been developed for hole circularity and hole taper. Optimal result for improved hole circularity and reduced hole taper has been obtained by particle swarm optimisation (PSO) technique. Effects of each input parameter on hole circularity and hole taper have been discussed and demonstrated graphically. Finally the experimental validation of the predicted results has been carried out.

2. EXPERIMENTAL WORK AND MEASUREMENT

2.1 Work Material

In present research, the Inconel-718 sheet of size 140x140x1.4 mm is taken for performing the required experiments. The Inconel-718 sheet is supplied by commercial vender –Randhir metal Pvt. Ltd. (Mumbai).

2.2 Experimental Setup

Laser drilling system consists of three major subsystems.

(a) Laser generation system,
(b) Laser delivery system,
(c) Work holding system

Present experiments are performed on solid state Pulsed Nd:YAG laser of 250W average power. Laser system have been design and developed by at Raja Rammna Center for Advanced Technology Indore, India (RRCAT-Indore). Laser drilling system is shown in Fig. 1 and system specification is given Table 1.

Table 1. Laser drilling system specification

Characteristics	Value
Average power	250 kW
Laser wavelength	1064 nm
Operation mode	Pulsed mode
Pulse width	10 nm
Transmission efficiency	90 %
Focused spot diameter	400 µm



Figure 1. Image of laser trepan drilling system.

2.3 Input Parameters Levels

The important controlling factors, assist gas pressure, laser current, stand–off distance, trepanning speed and pulse frequency have been considered for experiments. These parameters have been decided based on their importance or functional requirements for governing the geometrical quality characteristics.

2.3.1 Assist Gas Pressure

Function of assist gas is to eject the molten metal from drill area immediately after melting otherwise it get re-casted or re-solidify to parent metal and produces the poor quality hole. The pressure of assist gas governs the cooling and ejection of

metal properly. Higher pressure creates the force convection cooling and lower pressure may not be sufficient for proper ejection of metal.

2.3.2 *Laser Current*

The peak power is directly proportional to the lamp current as given by the Eqn. (1). Controlling the laser lamp current directly influences the peak power supplied to work material[21].

$$LaserPeakPower\,(P) = \eta k_0 I^{\gamma} \quad (1)$$

where η = conversion efficiency electrical lump in to laser output

k_o= constant for a given flash lamp

I = lamp current

2.3.3 *Stand-off Distance*

Stand-off distance (SOD) is the distance between nozzle tip to upper surface of work sheet. Changing the SOD changes the focal position of the laser beam. Laser power density is highest at the focal point of beam as focused dia is lowest as this point. If the focal plane is outside the material sheet then poorer melting and geometry will be resulted. Therefore, proper selection of SOD is required for better melting and quality.

2.3.4 *Trepanning Speed*

Trepanning speed is the relative speed between work sheet and laser delivery nozzle. Higher speed provides less time available for power supply at melting front and lower speed can be cause of uneven or over heating of drilling area.

2.3.5 *Pulse Frequency*

Laser beam is delivered in pulse mode. Energy supplied to work material depends on the number of beam pulses are being delivered in unit time. High frequency cause the large energy delivery and low frequency cause the poor melting of work sheet.

For finding the range of these input parameters, pilot experiments are performed for getting the through drilled holes. The five levels of input process parameters have been decided based on the observations of pilot experiments and machine capability. The different parameters and theirs values corresponding to the different levels are listed in Table 2.

Table 2. Levels of laser drilling process input parameters

Symbol	Input parameter	L-1	L-2	L-3	L-4	L-5
X_1	Assist gas pressure (bar)	6	7	8	9	10
X_2	Laser current (amp)	220	240	260	280	300
X_3	Stand-off distance (mm)	0.8	1.00	1.25	1.45	1.7
X_4	Trepanning speed (mm/ min)	10	20	30	40	50
X_5	Pulse frequency (Hz)	5	10	15	20	25

2.4 Measurements of Hole Diameters

Four diameter d_1, d_2, d_3, and d_4 are measured along the circumference of each drilled hole at interval of 45º angle. The measurements have been performed on tool bit type microscope having least count 0.001 cm and with maximum magnification of 10X. Figure 2 shows the scanned photograph of the specimen after laser drilling.

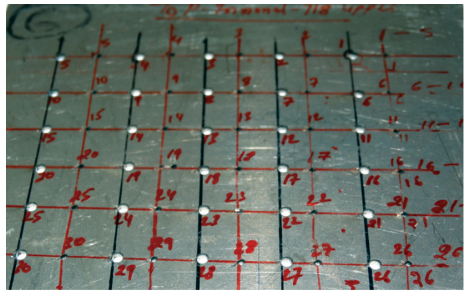


Figure 2. Scanned photograph of the specimen after drilling.

2.5 Calculation of Hole Quality Characteristics

The hole circularity has been defined as in Eqn. (2) and hole taper has been defined in Eqn. (3).

$$C_{ent} = \frac{(d_{\min})_{ent}}{(d_{\max})_{ent}} \quad (2)$$

$$\theta = \tan^{-1}\left(\frac{d_{ent} - d_{exit}}{2t}\right) X \frac{180}{\pi} \quad (3)$$

$$d_{ent} = \frac{(d_1 + d_2 + d_3 + d_4)}{4}$$

$$d_{exit} = \frac{(d_1 + d_2 + d_3 + d_4)}{4}$$

$(d_{min})_{ent}$= Minimum measured diameters at entrances side

$(d_{max})_{ent}$= Maximum measured diameters at entrances side

C_{ent} = circularity at entrance side

θ = Hole taper

The value of circularity and hole taper have been calculated for each drilled hole and graphically represented in Fig. 3. The Fig.4 indicates the variation of the measured diameters at entrance and exit during experimentation.

3. MATHEMATICAL MODELING OF QUALITY CHARACTERISTICS

With the help of regression analysis, a mathematical relation can be built between stated input variables and output features.

3.1 Regression Modelling

For better prediction of output parameter, the second order regression model is more suitable over first -order model[22] and general equation of second order regression model is given as Eqn. (4);

$$y = \beta_0 + \sum_{i=1}^{n}\beta_i X_i + \sum_{i=1}^{n}\beta_{ii} X^2{}_i + \sum_{i=1}^{n}\sum_{j=n+1}^{n}\beta_{ij} X_i X_j \quad (4)$$

β*s* are regression coefficients and value can be found by least square method, X_i (i= 1, 2, 3,- - - n) is value of n different input parameters, y is response parameter. Regression models

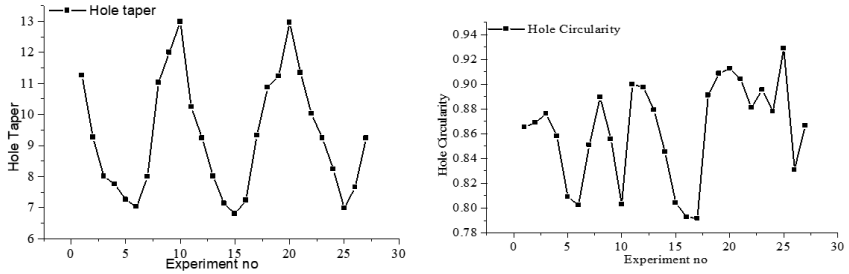


Figure 3. Numerical values of hole circularity and hole taper for each experiment.

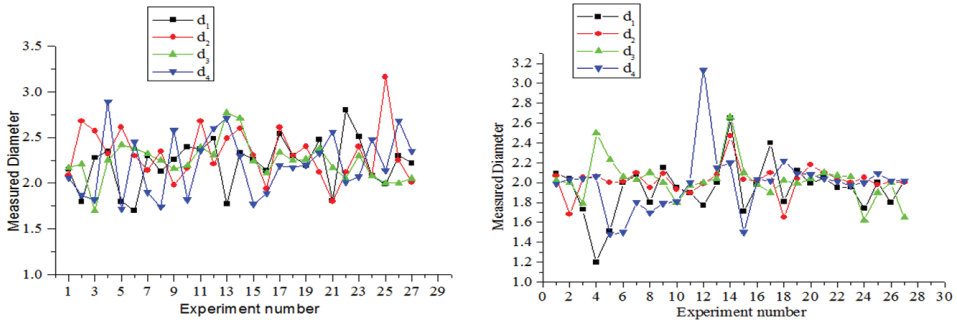


Figure 4. Variation of measured diameters at (a) entrance; (b) exit.

for hole circularity at entrance and hole taper are developed. The MINTAB14 software has been used to find the values of β and p-value of each term in regression models. Insignificant term in the model is removed by using back elimination process for better reliability of developed models. Final developed models for hole circularity and hole taper are Eqns. (5) and (6).

$$\begin{aligned} C_{ent} = {} & 0.502 + 0.0138 * X_1 + 0.00089 * X_2 \\ & + 0.0912 * X_3 - 0.00471 * X_4 + 0.0107 * X_5 \\ & - 0.0197 * X_1 * X1 - 0.000017 * X_2 * X_2 \\ & - 0.0783 * X_3 * X_3 + 0.000086 * X_4 * X_4 \\ & - 0.000440 * X_5 * X_5 + 0.00116 * X_1 * X_2 \end{aligned} \quad (5)$$

$$\begin{aligned} HT = {} & -16.2 + 7.80 * X_1 - 0.108 * X_2 + 2.56 * X_3 \\ & - 0.093 * X_4 - 0.150 * X_5 + 0.228 * X_1 * X_1 \\ & + 0.000908 * X_2 * X_2 - 2.55 * X_3 * X_3 \\ & - 0.00022 * X_4 * X_4 - 0.00144 * X_5 * X_5 \\ & - 0.0379 * X_1 * X_2 \end{aligned} \quad (6)$$

3.2 Model Validation and Adequacy Checking

Validity test has been carried out of developed regression models for that the S-values and coefficients of determination (R^2 and adjusted-R^2 values) are calculated for each model. The R^2 value for C_{ent} and HT models are 97.5 per cent and 84.0 per cent, respectively given in Table 3 which are in acceptable rages. Hence, the experimental data of C_{ent} and HT are well fitted in the developed models[23].

Analysis of variance has been carried out with the help of Minitab software as shown in Table 3. The p-values of the source of regression in both the models are lower than 0.01 and calculated F-ratios for source of regressions are 52.17, and 7.15 for C_{ent} and HT models respectively. The p-values and calculated F value as given in Table 3 and are well within range for making model significance[22,24-25]. Hence the developed models are significance and adequate for predicting the output response.

4. PARTICLE SWARM OPTIMISATION BASED OPTIMISATION

Particle swarm optimisation technique was introduced by Kennedy and Eberhart[18]. This technique is simple in concept, easy in implementation, and rapid convergence[26]. The PSO algorithm based on the facts that, each elements (birds) have own velocity and direction. Each bird communicates his position and direction to each of other birds; they identify the bird that is in the best location for reaching the destination, each bird, then, speeds up for best personal and best global (group) position in flocks. By doing so, each bird update their position and finds new local position. And this process continues until the flock reaches a desired destination.

PSO is computational economical in term of memory and speed required for computer system[27]. The flowchart of PSO

Table 3. ANOVA results for developed regression models

Response	Source	Degree of freedom	Sum of square	Mean of sum of square	F	P
C_{ent}	Regression	11	0.0409266	0.0037206	52.17	0.000
	Residual	15	0.0010697	0.0000713		
	Total	26	0.0419964			
HT	Regression	11	79.045	7.186	7.15	0.000
	Residual	15	15.072	1.005		
	Total	26	94.117			

Regression parameters for developed models

	S-Value	R^2 (%)	Adjusted-R^2 (%)
Circularity (C_{ent})	0.00844481	97.5	95.6
Hole taper (HT)	1.00240	84.0	72.3

algorithm is shown in Fig. 5[28]. The PSO algorithm consists of mainly five parts as follows.

(a) Problem definition
(b) Specifying the PSO parameters
(c) Initialisation of PSO
(d) Main loop of PSO
(e) Result of PSO

If fitness value is minimum value then optimum result obtained, otherwise PSO loop updates the position and velocity of each particle as per Eqns. (7) and (8). This process continues until best fitness value is obtained.

Objective of this research paper as mentioned earlier is to find maximum value of the hole circularity and minimum value of the hole taper in laser trepanned hole. Equation (6) of hole taper and Inverse of Eqn. (5) of hole circularity are separately defined in different M-file of MATLAB2008 software.

The main process parameters of PSO, like population size = 100, maximum iteration = 1000, Velocity inertia coefficient w = 0.025, Random number R_1 = 0.3, Random number R_2 = 0.3, C_1= 1.5 and C_2= 1.5 are taken for better performing PSO and as range suggested in literature available[29-31]. Number of variable in problem are S = 5.

After specifying the parameters, PSO algorithm initialised with randomly takes position value X_i (value of each variable in range taken is this paper). Value of initial particle velocity V_i can be randomly taken between 0 to 1.

Main loop of PSO start after initialisation, by using three parameters, current position X_i, personal best position P_i, and globally best position P_g. Each particle updates its velocity V_i to reach up the best particle g (solution). With the new velocity, each particle updates its position X_i as represented through Eqns. (7) and (8), respectively[27].

$$V_{inew} = wV_i + C_1R_1\,(P_i - X_i) + C_2R_2\,(P_g - X_i) \quad (7)$$

$$X_{inew} = X_i + V_{inew} \quad (8)$$

After defining the objective functions and specifying the PSO parameters, the developed MATLAB program have been executed and by reaching the termination criteria, the optimal value of input variables and optimal value (minimum) of inverse of circularity function and minimum value of hole taper function have been found. By taking inverse of PSO results minimum optimal value (1.03353) of 1/ C_{ent} gives the desired maximum values of hole circularity (C_{ent}). The optimal results obtained by the proposed methodology are given in Table 4. The optimal value suggested by PSO are better than all the calculated value of circularity and hole taper in experiments. The PSO resulted minimum value of inverse of circularity function and minimum of hole taper function are as shown in Fig. 6.

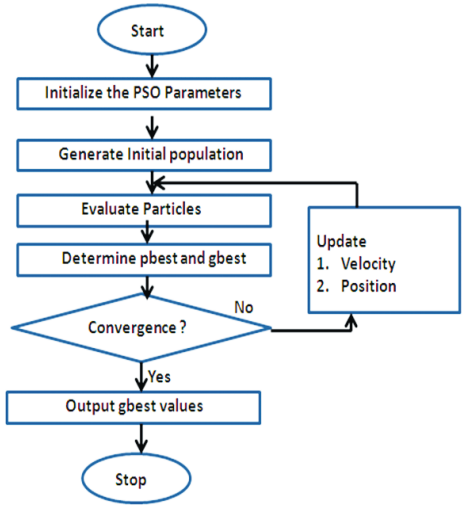


Figure 5 Flow chart of PSO algorithm[28].

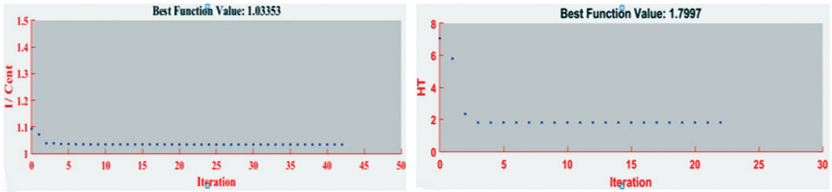


Figure 6. PSO results of minimum value of 1/ C_{ent} function and HT function.

5. RESULT DISCUSSION

5.1 Effect of Assist Gas Pressure and Laser Current on Quality Characteristics

The Fig. 7 shows the effect of assist gas pressure and laser current on the hole circularity and hole taper. By increasing the assist gas pressure, the hole circularity increases upto 8 bar after that it starts decreasing i.e. better circular holes may be created up to 8. Increasing the gas pressure up to 8 bar is responsible for the enhanced force available to eject the melted material and proper ejection of molten metal, minimizes the difference in diameters along the circumference of drilled hole. Due to that the circularity is increasing upto 8 bar. However, it is noted that if gas pressure is more than the required for proper ejection of molten metal, causes the force convection heat loss and rapid cooling of molten metal[10]. This resulting the widen the differences between diameters along the hole circumferences of drilled hole and hence reducing circularity by increasing the assist gas pressure after certain limit i.e. after 8 bar.

The molten material ejected from entrance side may recast at exit side of hole because of forced convection cooling of molten metal at higher gas pressure. Increasing mean entrance diameter and decreasing mean diameter by recasting at exit side, widen the difference between mean entrance and mean exit diameter. Hence increasing gas pressure, increases the hole taper.

As indicated in Fig. 7, by increasing the laser current, increase in the hole circularity has been observed at higher gas pressure but at lower assist gas pressure, no significant increment in circularity have been observed. At laser current 260 amp, combing with large gas pressure, resulting proper melting and ejection of work metal gives better circularity. The excess gas pressure after 8 bar promotes the forced convective cooling of molten metal and form the recast layer, which generates the poor circularity.

Hole taper increases with the increase in gas pressure because at the entrance side i.e. at the top surface of the specimen the molten material is easily ejected out from the cutting zone due to high gas pressure thereby offering more projected area for further melting of the material. However, at the bottom surface, the removal of the material is not proper, as such this leads to the solidification of molten material in the vicinity of the cutting zone. Thus, the drilled hole at the bottom surface is smaller in size as compared to the hole at the top surface. The hole taper is directly proportional to the difference in the top and bottom diameters. With the increasing gas pressure, the difference between these two diameters is enhanced thereby resulting in higher hole taper.

5.2 Effect of Pulse Frequency and Stand-off Distance on Quality Characteristics

By increasing pulse frequency, the hole circularity

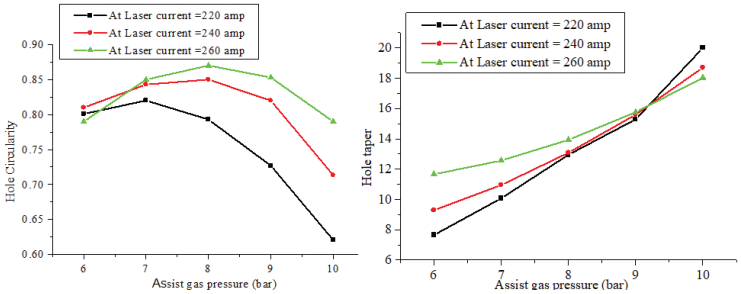


Figure 7. Effect of assist gas pressure and laser current on circularity and hole taper.

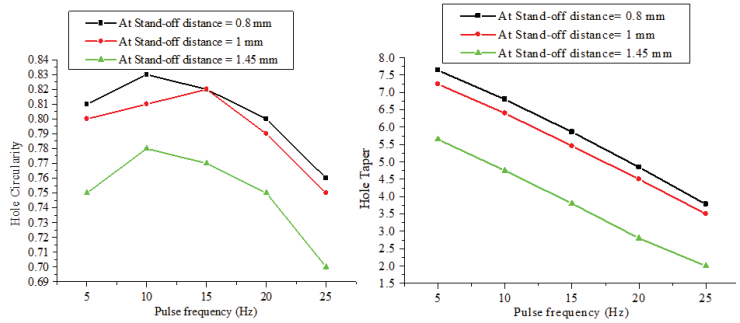


Figure 8. Effect of pulse frequency and SOD on circularity and hole taper.

decreases after 10 Hz in range taken. The hold value are Assist gas pressure = 6 bar, Laser current = 220 amp and Trepanning speed = 10 mm/min for the analysis. As shown in Fig. 8, At higher pulse frequency, laser energy is delivered repeatedly in short time, and molten material removal process may not finished before next pulse of energy delivered to drill area. That causes the excess heat reaches the material and increasing variation in diameters along the drill hole circumference is the reason for decreasing circularity.

Increasing SOD creates the positive focal plane position (plane at which laser beam is focused) in the sense that focal position is above the work surface. At positive focal position, work material receive less energy due to laser beam energy density is highest at focal plane compare to other plane[32]. Uneven melting of material produces variations in hole diameters along the drilled hole circumference. Therefore, the poor circularity has been observed. At higher pulse frequency combining with higher SOD, larger energy reaches to bottom surfaces, which narrowed the difference between mean entrance diameter and exit diameter. Hence reduces the hole taper.

5.3 Effect of Trepanning Speed and Pulse Frequency on Quality Characteristics

Decreasing pattern of circularity with increasing trepanning speed up to 30 mm/min is indicated in Fig. 9, there after increasing pattern has been observed with increasing trepanning speed. The hold values used for the analysis are Assist gas pressure = 6 bar, Laser current = 220 amp and SOD = 0.8 mm. At lower trepan speed, sufficient time is available for proper ejection of molten metal, and heat is restricted locally due to poor thermal conductivity of material. The molten material remains molten for relatively longer time and being ejected properly, therefore, the formation of recast layer is restricted. These phenomena are responsible for the better hole circularity at lower trepanning speed. However, by increasing the trepanning speed after 30 mm/min, restricts the overheating of material due to less time available for heating. The less time available also restrict the uneven heating and melting of material, and proper heating, melting and ejection of material make reducing variation in diameter measured along circumference. The reducing variation in diameter is the reason for increasing circularity.

At the higher speed combined with high pulse frequency, larger energy supplied for shorter period of time. Hence, the overheating and recast layer formation restricted at both side of hole and reduced hole taper can be obtained. At lower pulse frequency, less energy delivered to work material. At low energy, the random heating and melting makes poor circular hole. Similarly, at lower pulse frequency, the top side of hole melts more as compare to bottom side due less energy available and poor thermal conductivity of material. The large melting at top side widen the difference between mean entrance diameter and mean exit diameter. Widen difference in diameter is the reason for increasing hole taper with decreasing pulse frequency.

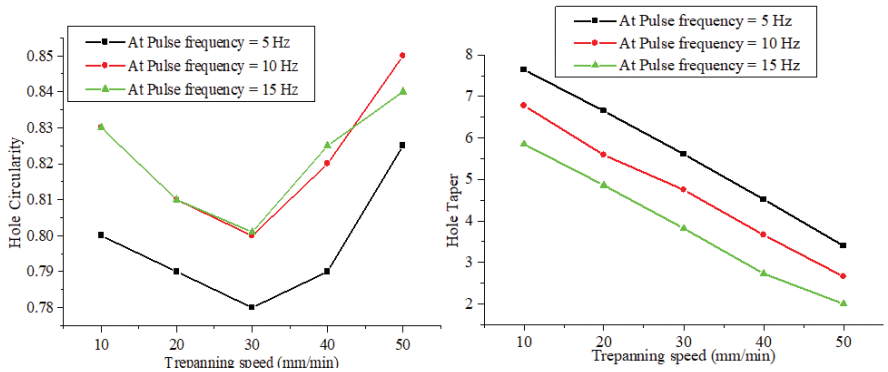


Figure 9. Effect of trepanning speed and pulse frequency on circularity and hole taper.

Table 4 Confirmation results

Response	Input parameter setting	Predicted values	Experimental values	Best value found during 27 experiments	Improvement (%)
Hole circularity	X_1= 9.20, X_2= 300, X_3= 0.80, X_4= 50, X_5= 12.20	0.97	0.98	0.93	5.37
Hole taper	X_1 = 6, X_2 = 220, X_3 = 0.8, X_4 = 10, X_5 = 5	1.80	2.85	6.65	57.14

5.4 Optimum Results

The optimal value of hole circularity has been found is 0.97, at level of assist gas pressure 9.20 bar, laser current 300 amp, SOD 0.8 mm, trepanning speed 50 mm/min and pulse frequency 12.20 Hz. For the experimental validation, the confirmation experiments have been conducted at optimum parameter setting. The comparison results are given in Table 4. From the Table 4, it is clear that the optimum experimental results show that an improvement of 5.37 per cent has been found in circularity as compared to the maximum value of hole circularity obtained during the experiments.

High laser current (300 amp) is responsible for high energy delivered to work material, but with combining with high trepanning speed (50 mm/min), high assist gas pressure (9.20), Lower SOD and medium pulse frequency provides the opportunity for proper heating and melting of work material. The higher Assist gas pressure removes all the melted material immediately from drilling area and high trepanning speed provides the shorter time available for laser beam interaction with work material, which makes better circular hole with improved circularity (0.97).

Similarly optimal value of Hole taper (HT) has been found is 1.8° at level of assist gas pressure 6 bar, laser current 220 amp, SOD 0.8 mm, trepanning speed 10 mm/min. and pulse frequency 5 Hz. For the experimental validation, the confirmation experiments have been conducted at optimum parameter setting. From the Table 4, it is clear that the optimum experimental results show that an improvement of 57.14 per cent has been found in hole taper as compared to minimum values obtained during the experimentation for the research i.e. out of 27 experiments.

Combing effects of low laser current (220 amp) and low frequency (5 Hz) is lower energy delivered to work material and due poor thermal conductivity of material, heat given to material confined in narrowed drilling area. Lower laser current with combination of low assist gas pressure (6 bar) and lower trepanning speed (10 mm/min.) limits the melting and removal of material from drilling area. Restricted melting and poor removal of molten metal generates the smaller diameter hole at entrance side. Lower entrance mean diameter narrows the difference between mean entrance diameter and mean exit diameter, which ultimately resulting in reduced hole taper (1.8°).

6. CONCLUSIONS

Laser trepanning drilling on material Inconel-718 has been performed for getting improved circularity and reduced hole taper. Following are the main finding of this research.

(a) Experiments of laser trepan drilling are performed on Inconel-718 extensively used in advanced mechanical component manufacturing.
(b) Reliable and adequate multi–regression mathematical models of hole circularity and hole taper are developed for predicting the hole quality characteristics.
(c) Computational intelligence technique, particle swarm optimisation (PSO) has been successfully applied for finding the optimal laser input parameters for improved circularity and reduced hole taper.
(d) Maximum circularity has been obtained at higher gas pressure, highest laser current, lowest SOD, highest trepan speed and medium pulse frequency.
(e) Reduced hole taper has been found at lower assist gas pressure, lower laser current, lower trepanning speed and low pulse frequency.
(f) The confirmation results show an improvement of 5.37 per cent and 57.14 per cent in hole circularity and hole taper at optimum parameter settings.

REFERENCES

1. Jain, V.K. Advanced machining processes. Allied publisher pvt., New Delhi, 2005. 2 -8 p.
2. Chyryssouris, G. Laser machining-Theory and practice. Springer, New york, 2005. 3 p.
3. Ezugwu, E.O.; Wang, Z.M. & Machado, A.R. The machinability of nickel based alloys - A review. *J. Mater. Proces. Technol.*, 1999, **86**, 1–16. doi: 10.1016/S0924-0136(98)00314-8
4. Yeo, C.Y.; Tam, Jana, S.C.S. & Lau, M.W.S. A technical review of the laser drilling of aerospace materials. *J. Mater. Proces. Technol.*, 1994, **42**, 15–49. doi: 10.1016/0924-0136(94)90073-6
5. Dubey, A.K. & Yadava, V. Laser beam machining – A review. *Int. J. Mach. Tools Manufacture*, 2008, **48**, 609-628. doi: 10.1016/j.ijmachtools.2007.10.017
6. Kumar, S.; Dubey, A.K. & Pandey, A.K. Computer-aided genetic algorithm based multi-opbjective optimization of laser trapan drilling. *Int. J. Pricision Eng. Manufacuring*, 2013, **14**, 1119-1125. doi: 10.1007/s12541-013-0152-5.
7. Choudhury, I. A.; Chong, W.C. & Vahid, G. Hole qualities in laser trepanning of polymeric materials. *Optics Lasers Eng.*, 2012 , **50**, 1297–1305. doi: 10.1016/j.optlaseng.2012.02.017
8. Pandey, A.K. & Dubey, A.K. Taguchi based fuzzy logic optimization of multiple quality characteristics in laser cutting of Duralumin sheet. *Optics Lasers Eng.*, 2012, **50** 328–335. doi: 10.1016/j.optlaseng.2011.11.005
9. Norkey, G.; Dubey, A.K. & Agarwal, S. Artficial intellegenec based modeling and optimization of heat affcted zone in Nd:YAG laser cutting of duralumin sheet. *J. Intel. Fuzzy Sys.*, 2014, **27**(3), 1545-1555. doi: 10.3233/IFS-141121
10. Chien, W.T. & Hou, S.C. Investigating the recast layer formed during the laser trepan drilling of Inconel 718 using the Taguchi method. *Int. J. Adv. Manufacturing Technol.*, 2007, **33,** 308–316. doi: 10.1007/s00170-006-0454-1
11. Goyal, R. & Dubey, A.K. Modeling and optimization of geometrical characteristics in laser trepan drilling of titanium alloy. *J. Mech. Sci. Technol.*, 2016, **30**(3) , 1281-1293. doi: 10.1007/s12206-016-0233-3
12. Marimuthu, S.; Antar, M.; Dunleavey, J.; Chantzis, D.; Darlington, W. & Hayward, P. An experimental study on quasi-CW fibre laser drilling of nickel superalloy. *Optics*

Laser Technol., 2017, **94**, 119–127. doi: 10.1016/j.optlastec.2017.03.021
13. Tamrin, K.F.; Nukman, Y.; Choudhury, I.A. & Shirley, S. Multiple-objective optimization in precision laser cutting of different thermoplastics. *Optics Lasers Eng.*, 2015, **67**, 57–65. doi: 10.1016/j.optlaseng.2014.11.001
14. Hajdarevic, D.B. & Bijelonja, I. Experimental and numerical investigation of temperature distribution and hole geometry during laser drilling process. *Procedia Engineering*, 2015, **100**, 384 – 393. doi: 10.1016/j.optlaseng.2014.11.001
15. Rajesh, P.; Nagaraju, U.; Harinath Gowd, G. & Vishnu Vardhan, T. Experimental and parametric studies of Nd:YAG laser drilling on austenitic stainless steel. *Int. J. Adv. Manufacturing Technol.*, 2017, **93**, 65–71. doi: 10.1007/s00170-015-7639-4
16. Chaudhary, A.K.; Pandey, A.K. & Dubey, A.K. Computer aided taguchi-fuzzymulti-optimization of laser cutting process. *J. Intel. Fuzzy Sys*., 2014, **26**(2), 801-810. doi: 10.3233/IFS-130770
17. Dhaker, K.L.; Pandey, A.K. & Upadhayay, B.N. Experimental investigation of hole diameter in laser trepan drilling of Inconel718 sheet. *Mater. Today: Proce.,* 2017, **4** , 7599–7608. doi: 10.1016/j.matpr.2017.07.093
18. Kennedy, J. & Eberhart, R.C. Particle swarm optimization. *In* Proceedings of IEEE International Conference on Neural Networks 2005: Piscataway, IEEE Service Center NJ, 2005, **5**, 1942–1948
19. Reyes-Sierra, M. & Coello, C. Multi-objective particle swarm optimizers: A Survey of the State-of-the-Art. *Int. J. Comput. Intel. Res.*, 2006, **2**(3) 287–308.
20. Adams II, Thomas A. & Seider, Warren D. Practical optimization of complex chemical processes with tight constraints. *Comput. Chem. Eng.,* 2008, **32**, 2099-2112. doi: 10.1016/j.compchemeng.2008.02.007
21. Joshi, P. & Shrma, A. Simultaneous optimization of kerf taper and heat affected zone in Nd-YAG laser cutting of Al 6061-T6 sheet using hybrid approach of grey relational analysis and fuzzy logic. *Precision Engineering*, 2018, **54**, 302-313. doi: 10.1016/j.precisioneng.2018.06.007
22. Montgomery, D.C. Design and analysis of experiments. wiley India, New Delhi, 2009. 389 p.
23. Phadke, M.S. Quality engineering using robust design. Prentice-hall, NJ, 1989. 51 p.
24. Gillett, B.E. Introduction to operation research –A computer–oriented algorithmic approach. Tata McGraw-Hill, New Delhi, 2008.
25. Pandey, A.K. & Dubey, A.K. Simultaneous optimization of multiple quality characteristics in laser cutting of titanium alloy sheet. *Optics Laser Technol.*, 2012, **44**, 1858–1865. doi: 10.1016/j.optlastec.2012.01.019
26. Mohanty, C.P.; Mahapatra, S.S. & Singh, M.R. A particle swarm approach for multi-objective optimization of electrical discharge machining process. *J. Intel. Manufacturing,* 2016, **27**, 1171–1190. doi: 10.1007/s10845-014-0942-3
27. Yusup, N.; Zain, A.M. & Mohd Hashim, S.Z. Overview of PSO for optimising process parameters of machining. *Procedia Engineering*, 2012, **29**, 914–923. doi: 10.1016/j.proeng.2012.01.064
28. Rao, R.V. Advanced modeling and optimisation of manufacturing process. Springer, London, 2011. 37 p.
29. Garg, H. & Sharma, S.P. Multi-objective reliability-redundancy allocation problem using particle swarm optimization. *Comp. Indust. Eng.*, 2013, **64**, 247–255. doi: 10.1016/j.cie.2012.09.015
30. Garg, S.; Patra, K. & Pal, S.K. Particle swarm optimization of a neural network model in a machining process. *Indian Academy Sci.*, 2014, **39**(3), 533–548. doi: 10.1007/s12046-014-0244-7
31. Das, S.; Abraham, A. & Konar, A. Particle swarm optimization and differential evolution algorithms: Technical analysis, applications and hybridization perspectives. *Stud. Comp. Intel.*, 2008, **116**, 1–38. doi: 10.1007/978-3-540-78297-1_1
32. Ghoreishi, M.; Low, D.K.Y. & Li, L. Compartive statistical analysis of hole taper and circularity in laser percussion drilling. *Int. J. Mach. Tool Manufacuring*, 2002, **42**, 985-995. doi: 10.1016/S0890-6955(02)00038-X

ACKNOWLEDGEMENT

We are thanks full to Dr B.N. Upadhayay, Senior scientist at Government of India's Raja Ramanna Centre for Advanced Technology (RRCAT)-Indore, India. Dr Upadhayay has supported us by allowing for conduction of experiments at their advanced laser cutting facility for this research purpose.

CONTRIBUTORS

Mr Kedari Lal Dhakar, has completed BE (Mechanical) from MITS-Gwalior (MP) and ME (Production Engineering) from Delhi College of Engineering, Delhi. Presently pursuing his PhD from Bundelkhand Institute of Engineering and Technology, Jhansi. Currently working as Assistant Professor in the Department of Mechanical Engineering at Jaypee University of Engineering & Technology, Guna.
He has conducted the experiments, measured the quality characteristics and analysed the experimental data by proposed methodology. He has also contributed in writing the research paper.

Dr Arun Kumar Pandey received his BTech (Mechanical Engineering) from Mahatma Gandhi Gramoday Vishwavidyalay Chitrakoot, Satna and MTech (Production) from M.I.T.S. Gwalior. He received his PhD from M.N.N.I.T. Allahabad. Presently working as Associate Professor at Bundelkhand Institute of Engineering and Technology, Jhansi. His research area is Advanced machining processes, Laser beam machining, application of artificial intelligence based tools for modeling and optimisation, application of design of experiment methods, laser cutting of difficult-to-laser cut materials, electrical discharge machining.
He has supervised to conduct the research and analysed the experimental data. He has also supervised the result and analysis and writing the research paper. He has also revised the manuscript based on the reviewers comments.

2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)

Performance Analysis of Different Bootstrap Inverter Configurations using Dual Gate OTFT

Neha Mishra[1], Poornima Mittal[2] and Brijesh Kumar[1]
Department of Electronics and Communication Engineering
Madan Mohan Malaviya University of Technology, Gorakhpur-273010 (Uttar Pradesh), India
Department of Electronics and Communication Engineering, Delhi Technological University, Delhi- 110042, India
nehamishra41193@gmail.com, poornima2822@ieee.org and bkece@mmmut.ac.in

***Abstract*: This paper presents the parameter extraction of various bootstrap techniques in terms of both static and dynamic performance. There are two configurations presented in the paper namely, Zero-V_{GS} load logic (ZVLL) and diode load logic (DLL). The performance is analyzed using three parameters –voltage gain, voltage transfer characteristics and dynamic response. This is an effective technique to overcome the threshold voltage drop in digital logic circuits. Comparison among these reveals that DLL configuration gives better dynamic results as compared to the ZVLL inverter. The paper concludes with the implementation of future issues to minimize the propagation delay and improve the noise margin. The comparison of performance in terms of static and dynamic responses of these three techniques is also dealt with.**

Keywords- stability, noise margin, inverter, transient

I INTRODUCTION

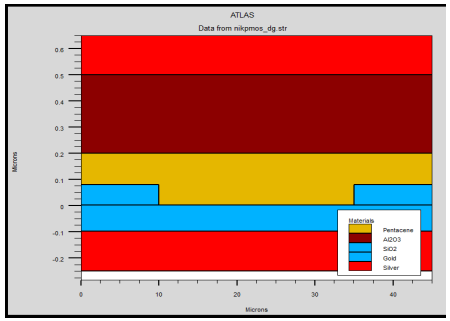
In the recent years the field of flexible electronics has survived remarkable developments. Organic semiconductors are inexpensive substitute to low cost large–area electronics. However, in conventional semiconductors the production cost and flexibility are obvious constraints. Universal logic gates and inverters are essential for applications of organic TFT's but only few logic designs have been reported to date [1-3].The performance of an organic inverter is affected by the process of fabrication. Uniformity in deposition of the layers is highly desirable for better performance of organic inverters. The different reasons for degradation in the transistor performance include its geometry defects, oxide breakdown, electro migration, hot carrier injection DIBL and ESD. These defects cause deflation in the performance of the transistor and the circuits. The external disturbances mentioned above include the noise, overvoltage and overcurrent that affect the circuit [10]. It will be possible in the upcoming years to use disposable organic circuits by printing it on a piece of paper, cloth; glass etc. because of its low cost fabrication processes [7-9]. Among two different types of structure based upon the position of gate relative to the organic semiconductor (OSC) are of two types - single gate and dual gate; dual gate configurations are superior in their performance as compared to single gate. Due to the presence of two gates it has higher charge injection area hence, better charge accumulation over single gate thin film transistor. The dual gate organic transistors (DG-OTFT) operate on the principle of accumulation rather than inversion as in the MOSFET devices. The charge transport in the transistors occurs due to the hopping mechanism. This mechanism explains about the movement of charge particles (holes and electrons) from HOMO and LUMO levels. For conduction to occur the holes (or electrons) should possess the energy which is higher than the band gap energy. In the organic semiconductor, proper overlapping of orbitals is also important. The commonly preferred overlapping among orbitals is sp^2-hybridized orbitals [2-4].
This paper is composed into six sections in which section 1 gives the introduction to the organic TFT and working principle whereas section 2and 3 explains the simulation setup and numerical iterations used in the extraction of parameters along with the dual gate based inverter configurations reported so far thereafter section 4 introduces the bootstrap technique and its different topologies further section 5 compares the performances of these topologies and hence, section 6 concludes the paper with the merits and demerits and scope of future enhancements.

II. SIMULATION SETUP

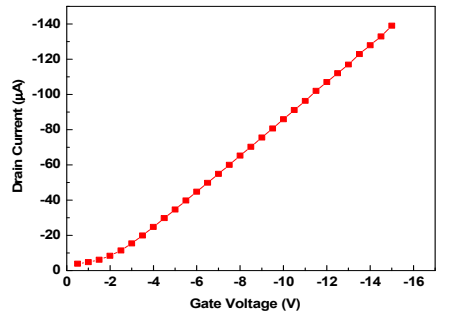
Dual gate based organic thin film transistor is studied using Silvaco atlas 2-D simulator. The simulated structure along with its input and output characteristics of device is shown in Fig. 1. The channel is formed with a p-type material which is Pentacene in this work. The mobility is analysed using Poole-Frenkel mobility model which is given as [9]:

$$\mu = \mu_o \left(\frac{\Delta}{kT} - 1\right) - \left(\frac{\beta}{kT} - \gamma\right) \qquad (1)$$

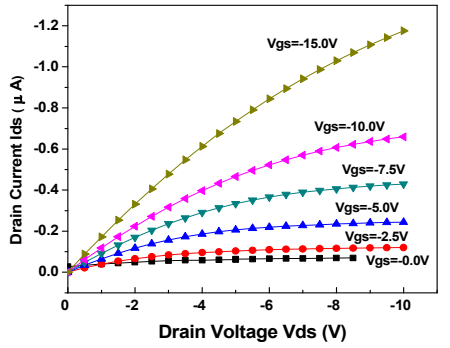
978-1-5386-5002-8/18/$31.00©2018IEEE



(a)



(b)



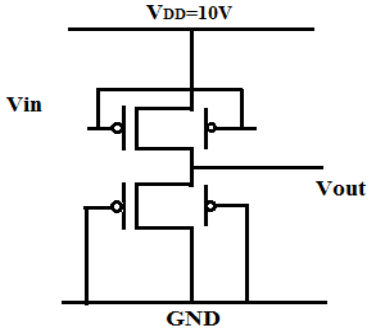
(c)

Fig. 1 (a) Simulated structure (b) input and output characteristics of dual gate based OTFT

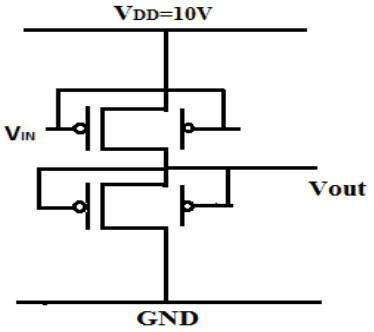
where μ is the field dependent mobility and μ_o is the zero field mobility and E is the electric field. Activation energy and Poole-Frenkel constant (for a p-type) are specified by Δ and β respectively. The values of Δ and β are given by 1.792 $*10^{-2}$ eV and 7.758 * 10^{-5} eV respectively [5-7]. The iterations are carried out using Newton Raphson (N-R) algorithm. The analysis is performed on both the steady state and transient state respectively. N-R algorithm is used under the mixed mode simulation for steady state analysis and transient state analysis. Moreover, in this work atlas full newton algorithm and modified newton algorithm are present for steady state simulation.

III. Dual Gate OTFT Based Inverter Configuration

There are two commonly used inverter configurations reported so far. They are zero-V_{GS} load logic (ZVLL) and diode load logic (DLL) [1].Both these configurations are suitable for specific applications. For faster switching applications DLL configuration is used and for better static performance ZVLL configuration is preferred. In the ZVLL configuration both the gate and source terminals are shorted resulting in a constant current source. As a result of which this structure it is always in ON state and load transistor is always in saturation and the driver transistor operates in the linear region [14-15].With these two basic inverter configurations different bootstrap techniques are implemented which are discussed in section 3. These techniques are used as a solution to the pitch falls in the circuit structure so as to enhance the performance both in respect to steady state and dynamic. The two inverter configurations are given in Fig. 2.



(a)

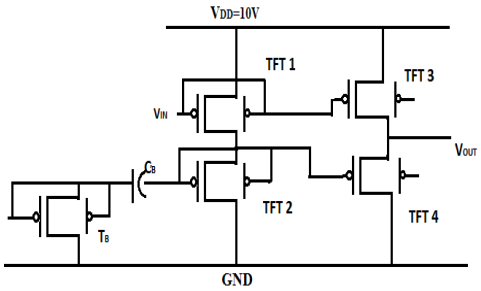


(b)

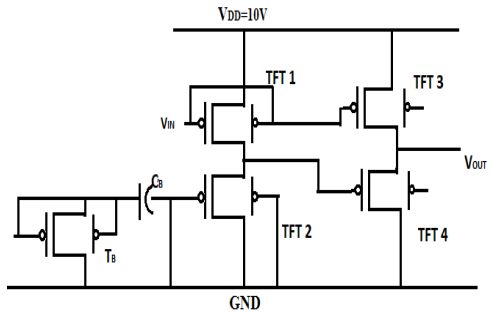
Fig. 2 Dual Gate OTFT based inverter configuration (a) DLL (b) ZVLL.

IV. Bootstrap Technique And Its Implementation In Dual Gate Inverter

To increase the robustness of the circuit in terms of noise margin, threshold voltage drop, voltage swing and gain, there is another technique known as bootstrap technique. Bootstrap technique is used to overcome the threshold voltage drop in the circuit and thereby enhance its performance [11-13]. The structure of the bootstrap technique employs a body transistor T_B and a capacitor C_B which is connected in the feedback. Ideally, the maximum pull up output voltage obtained by *p*-type transistor is V_{DD} but practically, it is limited to (V_{DD} - V_T). When this voltage drop will be minimized performance of the inverter will be enhanced.



(a)



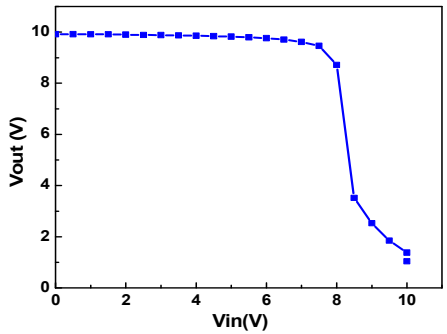
(b)

Fig. 3 Bootstrap inverter topologies using (a) ZVLL (b) DLL using DG-OTFT.

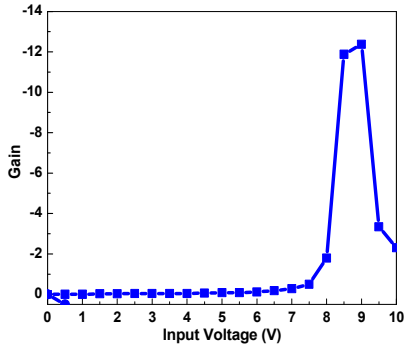
The feedback capacitor is used to minimize the voltage drop. The bootstrap technique in different dual gate based inverters is shown in Fig.3.

V. Extraction Performance Parameters of Different Bootstrap Topologies

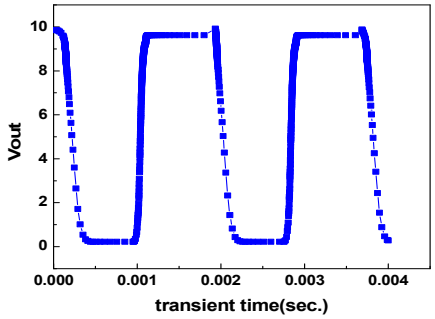
The main objective of the bootstrap technique is to improve the steady and dynamic performance of the circuit. Performances of the bootstrap topologies are compared in terms of voltage gain, output swing, VTC characteristics and transient characteristics. The VTC characteristics, voltage gain along with transient characteristics of different topologies are given in fig. 4, 5 and 6.



(a)

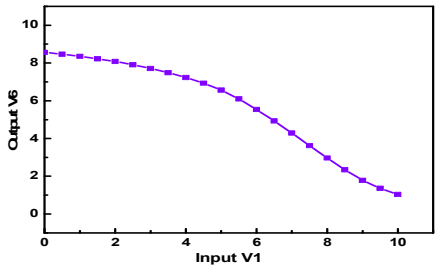


(b)

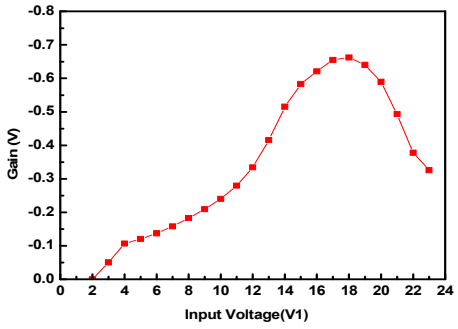


(c)

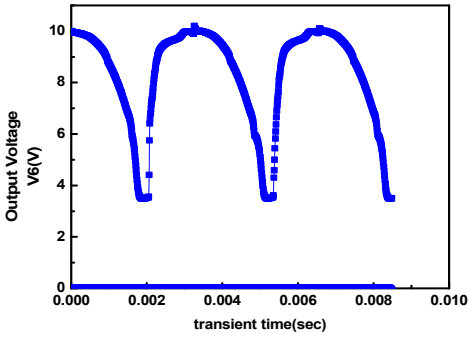
Fig. 4 Performance characteristics of DG-OTFT based bootstrap inverter in ZVLL configuration (a) VTC characteristics (b) voltage gain (c) Transient characteristics.



(a)



(b)



(c)

Fig. 5 Performance characteristics of DG-OTFT based bootstrap inverter in DLL configuration (a) VTC characteristics (b) voltage gain (c) Transient characteristics.

VI. Conclusion

The above paper presents the different bootstrap technologies with the objective to minimize the threshold voltage ratios in the logic circuits. From the above mentioned graphs it reveals that higher will be the value of output voltage better will be the static and dynamic performance of the circuit. In the designing of the bootstrap based circuit the key consideration is to increase the potential of the node so as to minimize the potential drop. These techniques are effective in designing of memory based circuits to increase its non-volatility and reduce the soft error. Different logic can be designed with the same objective to enhance the performance of the inverter.

REFERENCES

1. Y. Chen, D. Geng, M. Mativenga, H. Nam, and J. Jang, “High-speed pseudo-CMOS circuits using bulk accumulation a-IGZO TFTs,” IEEE Electron Device Lett., vol. 36, no. 2, pp. 153–155, Feb. 2014.
2. S. Negi, A. Rana, A. K. Baliga, P. Mittal and B. Kumar, ”Performance analysis of dual gate organic SR latch application”, *International Conference on Computing, Communication & Automation (ICCCA)*, 2015, Noida, India, pp. 1427 - 1432, May 15-16, 2015.
3. Kumar, B. K. Kaushik and Y. S. Negi, “Organic thin film transistors: Structure, models, materials, fabrication, and applications: A review*”, Polymer Review*, vol. 54, no. 1, pp. 33-111, Sept. 2013.
4. B. Kumar, B. K. Kaushik, Y. S. Negi, “Organic thin film [transistors: structures, models, materials, fabrication, and applications: A review”, *Polymer Reviews*, vol. 54. no.4, pp. 33–111, 2014.
5. B. Kumar, B.K. Kaushik, Y.S Negi, “Design and analysis of noise margin, write ability and read stability of organic and hybrid 6-T SRAM”, *Microelectron. Rel.* 54, vol. 12, pp. 2801–2812, 2014.
6. P. Mittal, B. Kumar, Y.S. Negi, B.K. Kaushik and R.K.Singh, “Channel length variation effect on performance parameters of organic field effect transistors,” *Microelectronics Journal*, vol. 43, pp. 985-994, 2012.
7. B. Kumar, B. K. Kaushik and Y. S. Negi, “Static and dynamic characteristics of dual gate organic TFT based NAND and NOR circuits*,” Journal of Computational Electronics*, vol. 13, no. 3, pp. 1-12, Sept. 2014.
8. Y. Chen, D. Geng, M. Mativenga, H. Nam, and J. Jang, “High-speed pseudo-CMOS circuits using bulk accumulation a-IGZO TFTs,” *IEEE Electron Device Lett.*, vol. 36, no. 2, pp. 153–155, Feb. 2014.
9. Shubham Negi, Yuvraj Singh Negi ,Vidhi Goswami “ Single and Dual gate OTFT based robust organic digital design.”, *Microelectronics Reliability,* pp.100-109,2014.
10. Q. Zhao, Y. Liu, J. Zhao, X. Guo, H. Li, and H. Yang, “Noise margin modeling for zero-VGS load TFT circuits and yield esti mation,”IEEE Trans.ElectronDevices,vol.63, no. 2, pp. 684–690, Feb. 2016.
11. B. Kumar, B.K. Kaushik, Y.S. Negi and V. Goswami, “Single and dual gate OTFTs based robust organic digital design”, *Microelectronic. Rel.*, vol. 54, no. 1, pp. 100–109, 2014.
12. P. Mittal, B. Kumar, Y.S. Negi, B.K. Kaushik and R.K.Singh, “Channel length variation effect on performance parameters of organic field effect transistors,” Microelectronics Journal, vol. 43, pp. 985-994, 2012.
13. Jong-Seok Kim,Jun-Hwan Jang, Yong-Duck Kim,Jung-Woo Byun,Kilim Han, Jin- Seong Park, and Byong-Deok Choi- “Dynamic Logic Circuits Using a-IGZO TFT’s, IEEE Transactions on Electron Devices, vol. 64, no. 10, pp. 4123-4130, Oct.2017.
14. Qing Wu, Jingyi Zhang and Qinru Qiu- “Design Considerations for Digital Circuits Using Organic Thin Film Transistors on a Flexible Substrate.” IEEE Symposium on Circuits and Systems, pp. 4, 2006.
15. Hirobumi Ushijima, Yasuyuki Kusaka, Mariko Fujita, Ken-ichi Nomura, Shusuke Kanazawa, Yoshinori Horii, Koji Abe and Noritaka Yamamoto- “ Developments of High Precision Printing Processes for Fabricating the Flexible Electronics”, ICEP, 2017.

International Journal of Research in Engineering and Innovation (IJREI)
journal home page: http://www.ijrei.com
ISSN (Online): 2456-6934

Performance improvement of vapour compression refrigeration systems

R.S. Mishra

Department of Mechanical & Production Engineering, Delhi Technological University Delhi, India

Abstract

The improvement of performance of system is too important for higher refrigerating effect or reduced power consumption for same refrigerating effect. Many efforts have to be done to improve the performance of vapour compression refrigeration system**.** For improving the coefficient of performance, it required reduced compressor work and increasing refrigeration effect. The decrease in condenser pressure and temperature, the refrigeration effect will increase and compressor input work also increases coefficient of performance (COP). To improve thermal performance of vapour compression refrigeration systems by improving first law efficiency (COP), second law efficiency (exergetic efficiency) by reducing of system defect in components of VCRS which results into reduction of work input through detailed analysis of vapour compression refrigeration systems using ecofriendly refrigerants. This paper mainly deals with the effect of various thermodynamic parameters (i.e. evaporator and condenser temperatures. effect of super heating in evaporator outlet and sub-cooling of condenser outlet, liquid vapour heat exchanger effectiveness, compressor efficiency) variation on first law efficiency (COP), second law efficiency (exergetic efficiency and rational exergy destruction ratio of vapour compression refrigeration system. Using energy-exergy analysis, it is found that thermodynamic performances of HFC-134a refrigerant is slightly higher than HFO refrigerants (i.e. R-1234ze & R-1234yf). Although thermal performances of HFO-1234ze is higher (around 5% to 10%) than HFO-1234yf. Therefore HFO-1234 refrigerants can replace HFC-134a refrigerant in vapour compression refrigerant without system modification.
© 2019 ijrei.com. All rights reserved

Keywords: Thermodynamic performance improvements, Mixing Nano materials, Vapour Compression Refrigeration System.

1. Introduction

Refrigeration is a technology which absorbs heat at low temperature and provides temperature below the surrounding by rejecting heat to the surrounding at higher temperature. Vapour compression system which consists of four major components compressor, expansion valve, condenser and evaporator in which total cooling load is carried at one temperature by single evaporator but in many applications like large hotels, food storage and food processing plants, food items are stored in different compartment and at different temperatures. Therefore there is need of improving performances of vapour compression refrigeration system. The systems under vapour compression technology consume huge amount of electricity, this problem can be solved by improving performance of system. Thermodynamic performances of systems based on vapour compression refrigeration technology can be improved by following:

(i) The performance of refrigerator is evaluated in term of COP which is the ratio of refrigeration effect to the net work input given to the system. The COP of vapour compression refrigeration system can be improved either by increasing refrigeration effect or by reducing work input given to the system.

(ii) It is well known that throttling process in VCR is an irreversible expansion process. Expansion process is one of the main factors responsible for exergy loss in cycle performance because of entering the portion of the refrigerant flashing to vapour in evaporator which will not only reduce the cooling capacity but also increase the size of evaporator. This problem can be eliminated by adopting multi-stage expansion with flash chamber where the flash vapour is removed after each stage of expansion as a consequence there will be increase in cooling capacity and reduce the size of the evaporator.

(iii) Work input can also be reduced by replacing multi-stage compression or compound compression with single stage compression.

Corresponding author: R. S. Mishra
Email Address: hod.mechanical.rsm@dtu.ac.in

(iv) The refrigeration effect can also be increased by passing the refrigerant through sub-cooler after condenser to evaporator.

(v) Use of Nano refrigerant in the primary circuit vapour compression refrigeration system /secondary circuit of evaporator improves thermodynamic performances.

(vi) Use of Nano refrigerant in the secondary circuit of evaporator /and secondary circuit of condenser improves thermodynamic performances

The vapour compression refrigeration system based applications make use of refrigerants which are responsible for greenhouse gases, global warming and ozone layer depletion. Montreal protocol was signed on the issue of substances that are responsible for depleting Ozone layer and discovered how much consumption and production of ozone depletion substances took place during certain time period for both developed and developing countries. Another protocol named as Kyoto aimed to control emission of greenhouse gases in 1997. The relationship between ozone depletion potential and global warming potential is the major concern in the field of GRT (green refrigeration technology) so Kyoto proposed new refrigerants having lower value of ODP and GWP. Internationally a program being pursued to phase out refrigerants having high chlorine content for the sake of global environmental problems .Due to presence of high chlorine content ,high global warming potential and ozone depletion potential after 90's CFC and HCFC refrigerants have been restricted. Thus, HFC refrigerants are used nowadays, showing much lower global warming potential value, but still high with respect to non-fluorine refrigerants. Lots of research work has been done for replacing "old" refrigerants with "new" refrigerants. [1]

2. Replacement of HFC-134a by HFO Ecofriendly Refrigerants

The European Union has approved a regulation (Regulation (EU) No 517/2014) which prohibits the use of fluorinated gases with a GWP more than 150 in domestic refrigerators and freezers w.e.f.1st Jan, 2015, and in new types of mobile air-conditioning systems from January 1, 2017. Therefore the low GWP and zero ODP alternative refrigerants R1234yf and R1234ze may be investigated as replacements of HFCs. The use of alternatives to HFCs in the refrigeration and air-conditioning systems controls environmental pollution as well as supports healthy nature and surroundings.

Esbir et al. [2] carried out experimental analysis of R1234yf as a drop-in replacement for R134a in a vapour compression system and compared the energy performance of both refrigerants, R134a and R1234yf, in a observed vapour compression system under a extensive range of working conditions. Zhao et al. [3] had developed mini-channel evaporator model using HFO1234yf as working fluid and proposed simulation model of the mini-channel evaporator finite conception and effectiveness by using NTU method for calculating the heat transfer rate and also compared experimental data of six different samples using R1234yf under typical working conditions of an automotive air conditioning system. Nikolaidis and Probert [4] analytically studied the change in evaporator and condenser temperatures of two stage vapour compression refrigeration plant using R22 and found additional considerable effect on plant irreversibility and suggested that there is need for optimizing the conditions imposed upon the condenser and evaporator. Getu and Bansal [5] optimized the design and operating parameters such as condensing temperature, subcooling temperature, evaporating temperature ,superheating temperature and temperature difference in cascade refrigeration system using R717 in high temperature heat exchanger (condenser) and R744 cascading in low temperature heat exchanger (Evaporator) using regression analysis to obtain optimum thermodynamic parameters of two stage cascade vapour compression refrigeration system. Spatz and Motta [6] experimentally investigated replacement of R12 with R410a through of medium temperature vapour compression refrigeration cycles in terms of heat transfer and pressure drop characteristics comparison using thermodynamic analysis and found that the R410a gives best performance among R12, R404a and R290a. Mohanraj et al. [7] carried out experimental investigations on domestic refrigerator and found that under different environmental temperatures, the COP of system using mixture of R290 and R600a in the ratio of 45.2: 54.8 by weight showing up to 3.6% gives better performance than same system using R134a because discharge temperature of compressor with mixture of R290 and R600a is lower than same compressor with R134a. Mastani Joybari et al. [8] carried out experimental investigations on a domestic refrigerator using 145g R134a and concluded that the exergetic defect occurred in compressor was maximum as compare to other components and also found that instead of 145g of R134a if 60g of R600a is used in the considered system gave same performance which had economical advantages in terms of reduced risk of flammability of hydrocarbon refrigerants. Han et al. [9] found mixture of R32/R125/R161 showing higher COP, less pressure ratio and slightly high discharge compressor temperature without any modification in the same system under different working conditions could be replacement of R407C in vapour compression refrigeration system having rotor compressor by conducting experimental measurements. Halimic et al. [10] compared thermodynamic performance of vapour compression refrigeration system using R401A, R290 and R134A with R12 and found similar performance of R134a in comparison with R12,R134A in the same system without any medication in the system components and concluded that hydrocarbon (R290) gives top results in reference to greenhouse impact. Ahamed et al. [11] mainly emphasized on use of hydrocarbons and mixture of hydrocarbons and R134a in vapour compression refrigeration system and found that the compressor shows much higher exergy destruction as compared to other components of vapour compression refrigeration system and concluded that this exergy destruction can be minimized by using of nanofluid and nanolubricants in compressor. Ahamed

et al. [12-13] had performed theoretical & experimental investigations of domestic refrigeration system using hydrocarbons (isobutene and butane) using energy and exergy analysis and found that the energy efficiency ratio of hydrocarbons is comparable with R134a. The exergy efficiency and sustainability index of hydrocarbons much higher than that of R134a at considered evaporator temperature and also found that the compressor shows highest system defect (69%) among other components of considered system. Reddy et al. [14] carried out numerical analysis of vapour compression refrigeration system using R134a, R143a, R152a, R404A, R410A, R502 and R507A to find out the the effect of evaporator temperature, degree of subcooling at condenser outlet, superheating of evaporator outlet, vapour liquid heat exchanger effectiveness and degree of condenser temperature on COP and exergetic efficiency and concluded that the evaporator and condenser temperatures have important effect on both COP and exergetic efficiency and also found that R134a has the better performance. Although the R407C has poor performance in all respect. Saravana kumar and Selladurai [15] compared the performance using R134a and R290/R600a mixture on a domestic refrigerator and found that the hydrocarbon mixture of R290/R600a showed advanced COP and exergetic efficiency than R134a and found maximum irreversibility in the compressor as compared to condenser, expansion valve and evaporator. Kumar et al. [16] carried out energy and exergy analysis for calculating the coefficient of performance and exergetic efficiency, various losses occurred in different components of vapour compression cycle using R11 and R12 as refrigerants by the use of exergy-enthalpy diagram. Anand and Tyagi [17] carried out detailed exergy analysis of window air conditioning test rig with R22 as working fluid of two ton of refrigeration capacity and concluded that the irreversibility in system components will be highest when the system is 100% charged and lowest when 25% charged andfound that the irreversibility in compressor is maximum among other system components. Xuan and Chen [18]presented in this manuscript about the replacement of R502 by mixture of HFC-161.Through experimental study it was found that mixture of HFC-161 gives same and higher performance than R404A at lower and higher evaporative temperature respectively on the vapour compression refrigeration system designed for R404A. Cabello et al. [19]studied the outcome of operating parameters on COP, work input and cooling capacity of single-stage vapour compression refrigeration system and found the excessive influence on constraints in energetic performance due change in suction pressure, condensing and evaporating temperatures. Cabello et al. [20]experimentally investigated the effect of condensing pressure, evaporating pressure and degree of superheating on single stage vapour compression refrigeration system using R22, R134a and R407C and found that the mass flow rate is greatly affected by change in suction conditions of compressor in results on refrigeration capacity because refrigeration capacity depended on mass flow rate through evaporator and also concluded that for higher compression ratio R22 gives lower COP than R407C. Stanciu et al. [21] carried out numerical and graphical investigation in terms of COP, compressor work, exergy efficiency and refrigeration effect on single stage vapour compression refrigeration system using R22, R134a, R717, R507a, R404a refrigerants. The effect of sub-cooling, superheating and compression ratio was also studied by them on the same system using considered refrigerants and presented system optimization of vapour compression refrigeration working with the specific refrigerant. Bolaji et al. [22] carried out experimentally comparative analysis of R32, R152a and R134a refrigerants in vapour compression refrigerator and concluded that the R32 had lowest thermodynamic performance than R134a and R152a. However R134a and R152a showing nearly similar performance but greatest performance was obtained of VCRS system using R152a. Yumrutas et al. [23]carried out exergy analysis on vapour compression refrigeration cycle in terms of pressure losses, COP, second law efficiency and exergy losses based investigation of effect of condensing and evaporating temperature and found that the first law efficiency in terms of Coefficient of performance (COP) increases with increase in evaporator and exergy efficiency decreases . The total exergy losses of system increases with condenser temperature and evaporator temperatures. Padilla et al. [24] carried out exergy analysis for replacing R12 by R134a of domestic vapour compression refrigeration system using R12 and R R134a refrigerants and concluded that the performance in terms of power consumption, irreversibility and exergetic efficiency of R134a is slightly higher than R12. Arora and Kaushik [25] developed numerical model of actual vapour compression refrigeration system with liquid vapour heat exchanger and carried out energy and exergy analysis on the same system in the specific temperature range of evaporator and condenser and concluded that the R502 is the superlative refrigerant as compared to R404A and R507A and found that the compressor is the worst and liquid vapour heat exchanger is best component of the vapour compression refrigeration system. Based on the literature it was observed that researchers have gone through detailed first law analysis in terms of coefficient of performance and second law analysis in term of exergetic efficiency of simple vapour compression refrigeration system with single evaporator. Researchers did not go through the irreversibility analysis or second law analysis of simple compression in vapour compression refrigeration systems. To improve thermal performance of vapour compression refrigeration systems by improving first law efficiency (COP) , second law efficiency (exergetic efficiency) by reducing of system defect in components of VCRS which results into reduction of work input through detailed analysis of vapour compression refrigeration systems using ecofriendly refrigerants.

3. Result and Discussion

Table-1(a) shows the variation of condenser temperature with the variation of first law efficiency in terms of coefficient of performance (COP) of HFO refrigerants and HFC-134a refrigerant. The first law efficiency in terms of COP is decreasing as condenser temperature is increasing. It is clear that COP of HFC-134a is higher than HFO refrigerants. Although COP of R1234ze is similar to R134a but higher than R1234yf. Table-1(b) shows the condenser temperature variation with rational exergy destruction ratio ($EDR_{_Rational}$) of HFO refrigerants and HFC-134a. The condenser temperature is increasing as rational exergy destruction ratio ($EDR_{_Rational}$) is also increasing. It is clear that rational exergy destruction ratio ($EDR_{_Rational}$) of HFC-134a is lower than HFO refrigerants. Although rational exergy destruction ratio ($EDR_{_Rational}$) of R1234ze is slightly higher than to R134a. The highest ($EDR_{_Rational}$) is found by using R1234yf as compared to HFC -134a. Table-1(c) shows the variation of condenser temperature variation of second law efficiency (exergetic efficiency) HFO refrigerants and HFC-134a. The second law efficiency (exergetic efficiency) is decreasing as condenser temperature is increasing. It is clear that exergetic efficiency of HFC-134a is higher than HFO refrigerants. Although exergetic efficiency of R1234ze is slightly lower than R134a but higher than R1234yf.

Table-1(a) Comparison of first law efficiency (in terms of Coefficient of performance) of HFO refrigerants with HFC-123a in the vapour compression refrigeration using liquid vapour heat exchanger using R134a ecofriendly refrigerant. (For T_eva=253K, Compressor efficiency=0.8, Heat exchanger effectiveness = 0.80, T_super heat_Eva=10, T_sub_cool_Cond=10

T_Cond (K)	COP_R134a	COP_R1234ze	COP_R1234yf
333	1.922	1.889	1.755
328	2.168	2.141	2.015
323	2.449	2.428	2.310
318	2.774	2.761	2.649
313	3.159	3.154	3.047
308	3.622	3.827	3.524
303	4.194	4.211	4.111

Table-1(b) Comparison of exergy Destruction Ratio of HFO refrigerants with HFC-123a in the vapour compression refrigeration using liquid vapour heat exchanger using R1234ze ecofriendly refrigerant. (For T_eva=253K, Compressor efficiency=0.8, Heat exchanger effectiveness=0.80, T_superheat_Eva=10, T_sub_cool_Cond=10

T_Cond (K)	EDR_ Rational R134a	EDR_Rational R1234ze	EDR_Rational R1234yf
333	0.6581	0.6640	0.6878
328	0.6143	0.6192	0.6415
323	0.5644	0.5681	0.5891
318	0.5065	0.5089	0.5288
313	0.4382	0.4390	0.4580
308	0.3558	0.3549	0.3732
303	0.2540	0.251	0.2689

Table-1(c) Comparison of second law efficiency (in terms of exergetic efficiency) of HFO refrigerants with HFC-123a in the vapour compression refrigeration using liquid vapour heat exchanger using R1234yf ecofriendly refrigerant. (For T_eva=253K, Compressor efficiency=0.8, Heat exchanger effectiveness=0.80, T_superheat_Eva=10, T_sub_cool_Cond=10

T_Cond (K)	Exergetic Eff_ R134a	Exergetic Eff_ R1234ze	Exergetic Eff _R1234yf
333	0.3419	0.3360	0.3122
328	0.3857	0.3808	0.3585
323	0.4356	0.4319	0.4109
318	0.4935	0.4911	0.4712
313	0.5618	0.5610	0.5420
308	0.6442	0.6451	0.6268
303	0.7460	0.7490	0.7311

Table-2(a) shows the variation of evaporator temperature variation of first law efficiency in terms of coefficient of performance (COP) of HFO refrigerants and HFC-134a. The first law efficiency in terms of COP is increasing as evaporator temperature is increasing. It is clear that COP of HFC-134a is higher than HFO refrigerants. Although COP of R1234ze is similar to R134a but higher than R1234yf.

Table-2(b) shows the evaporator temperature variation with rational exergy destruction ratio ($EDR_{_Rational}$) of HFO refrigerants and HFC-134a. The evaporator temperature is increasing as rational exergy destruction ratio ($EDR_{_Rational}$) is also increasing. It is clear that rational exergy destruction ratio ($EDR_{_Rational}$) of HFC-134a is lower than HFO refrigerants. The high ($EDR_{_Rational}$) is found by using R1234yf as compared to HFC-134a and HFO-1234ze refrigerants. Although rational exergy destruction ratio ($EDR_{_Rational}$) of R1234ze is slightly higher than to R134a.

Table-2(c) shows the variation of evaporator temperature variation of second law efficiency (exergetic efficiency) HFO refrigerants and HFC-134a. The second law efficiency (exergetic efficiency) is decreasing as evaporator temperature is increasing. It is clear that second law efficiency of HFC-134a is higher than HFO refrigerants. Although second law efficiency of R1234ze is similar to R134a but higher than R1234yf.

Table-2(a) Variation of evaporator temperature with thermal performance of vapour compression refrigeration using liquid vapour heat exchanger using R134a ecofriendly refrigerant. (For T_cond=323K , Compressor efficiency=0.8, Heat exchanger effectiveness=0.80, T_superheat_Eva=10, T_sub_cool_Cond=10

$T_{_EVA}$ (K)	COP_actual R134a	COP_Actual R1234ze	COP_ R1234yf
253	2.449	2.428	2.31
258	2.796	2.783	2.656
263	3.215	3.21	3.075
268	3.73	3.736	3.59
273	4.376	4.396	4.237
278	5.21	5.246	5.072
283	6.325	6.384	6.191

R. S Mishra / International journal of research in engineering and innovation (IJREI), vol 3, issue 1 (2019), 48-56

Table-2(b) Comparison of exergy Destruction Ratio of HFO refrigerants with HFC-123a in the vapour compression refrigeration using liquid vapour heat exchanger using R1234ze ecofriendly refrigerant. (For T_cond=323K, Compressor efficiency=0.8, Heat exchanger effectiveness=0.80,T_superheat_Eva=10, T_sub_cool_Cond=10

T_{EVA} (K)	EDR_Rational R134a	EDR_Rational R1234ze	EDR_Rational R1234yf
253	0.4490	0.5681	0.5891
258	0.5665	0.5686	0.5882
263	0.5721	0.5728	0.5808
268	0.5824	0.5818	0.5982
273	0.5992	0.5975	0.6120
278	0.6252	0.6226	0.6351
283	0.6647	0.6616	0.6719

Table-2(c) Comparison of second law efficiency (in terms of exergetic efficiency) of HFO refrigerants with HFC-134a in the. vapour compression refrigeration using liquid vapour heat exchanger (For T_cond=323K, Compressor efficiency=0.8, Heat exchanger effectiveness=0.80, T_superheat_Eva=10, T_sub_cool_Cond=10

T_{EVA} (K)	Exergetic Efficiency	Exergetic Efficiency	Exergetic Efficiency
253	0.4356	0.4319	0.4109
258	0.4335	0.4314	0.4118
263	0.4279	0.4272	0.4092
268	0.4176	0.4182	0.4092
273	0.4008	0.4025	0.4018
278	0.3748	0.3774	0.388
283	0.3353	0.3334	0.3649

Table-3(a) shows the effect of super heating temperature variation of HFO refrigerants and HFC-134a. The first law efficiency in terms of COP is increasing as super heating temperature is increasing. It is clear that COP of HFC-134a is higher than HFO refrigerants. Although COP of R1234ze is similar to R134a but higher than R1234yf.

Table-3(b) shows the super heating temperature variation with rational exergy destruction ratio ($EDR_{_Rational}$) of HFO refrigerants and HFC-134a. The super heating temperature is increasing as rational exergy destruction ratio ($EDR_{_Rational}$) is also decreasing. It is clear that rational exergy destruction ratio ($EDR_{_Rational}$) of HFC-134a is lower than HFO refrigerants. The high ($EDR_{_Rational}$) is found by using R1234yf as compared to HFC-134a and HFO-1234ze refrigerants. Although rational exergy destruction ratio ($EDR_{_Rational}$) of R1234ze is slightly higher than to R134a. Table-3(c) shows the variation of super heating temperature variation of second law efficiency (exergetic efficiency) HFO refrigerants and HFC-134a. The second law efficiency (exergetic efficiency) is increasing as super heating temperature is increasing. It is clear that second law efficiency of HFC-134a is higher than HFO refrigerants. Although second law efficiency of R1234ze is similar to R134a but higher than R1234yf.

Table-3(a) Comparison of First law efficiency (in terms of COP) of HFO refrigerants with HFC-134a in the vapour compression refrigeration using liquid vapour heat exchanger using ecofriendly refrigerants. (For T_eva=253K, T_Cond=323K, Compressor efficiency=0.8, Heat exchanger effectiveness=0.80, T_sub_cool_Cond=10

Super heating Temperature	COP_Actual R134a	COP_Actual R1234ze	COP_Actual R1234yf
0	2.448	2.407	2.278
5	2.449	2.417	2.293
10	2.450	2.428	2.310
15	2.451	2.44	2.328
20	2.452	2.452	2.347

Table-3(b) Effect of super heating of evaporator on comparison of exergey Destruction Ratio of vapour compression refrigeration using liquid vapour heat exchanger using ecofriendly refrigerants. (For T_eva=253K, T_Cond=323K, Compressor efficiency=0.8, Heat exchanger effectiveness=0.80, T_superheat_Eva=10, T_sub_cool_Cond=10

Super heating Temperature	EDR_ Rational R134a	EDR_ Rational R1234ze	EDR_ Rational R1234yf
0	0.5652	0.5718	0.5948
5	0.5651	0.5701	0.5922
10	0.5650	0.5681	0.5891
15	0.5649	0.5660	0.5859
20	0.5648	0.5648	0.5826

Table-3(c) Effect of super heating of evaporator on second law efficiency (in terms of exergetic efficiency) of HFO refrigerants with HFC-134a in the thermal performance of vapour compression refrigeration using liquid vapour heat exchanger using R1234yf ecofriendly refrigerant. (For T_eva=253K, T_Cond=323K Compressor efficiency=0.8, Heat exchanger effectiveness=0.80, ,T_sub_cool_Cond=10

Super heating Temperature	Exergetic Efficienc R134a	Exergetic Efficiency R1234ze	Exergetic Efficiency R1234yf
0	0.4358	0.4282	0.4052
5	0.4359	0.4299	0.4078
10	0.4360	0.4319	0.4109
15	0.4361	0.4340	0.4141
20	0.4362	0.4362	0.4174

Table-4(a) shows the Effect of subcooling temperature variation with first law efficiency (COP) of HFO refrigerants and HFC-134a. The first law efficiency in terms of COP is increasing as subcooling temperature is increasing. It is clear that COP of HFC-134a is higher than HFO refrigerants. Although COP of R1234ze is similar to R134a but higher than R1234yf. Table-4(b) shows the sub cooling temperature variation with rational exergy destruction ratio ($EDR_{_Rational}$) of HFO refrigerants and HFC-134a. The sub cooling temperature is increasing as rational exergy destruction ratio ($EDR_{_Rational}$) is also decreasing. It is clear that rational exergy destruction ratio ($EDR_{_Rational}$) of HFC-134a is lower than HFO refrigerants. The high ($EDR_{_Rational}$) is found by using R1234yf

as compared to HFC-134a and HFO-1234ze refrigerants. Although rational exergy destruction ratio ($EDR_{_Rational}$) of R1234ze is slightly higher than to R134a. Table-4(c) shows the variation of sub cooling temperature variation with second law efficiency (exergetic efficiency) HFO refrigerants and HFC-134a. The second law efficiency (exergetic efficiency) is increasing as sub cooling temperature is increasing. It is clear that second law efficiency (exergetic efficiency) of HFC-134a is higher than HFO refrigerants. Although second law efficiency (exergetic efficiency) of R1234ze is slightly lower to R134a but higher than R1234yf.

Table-4(a) Effect of subcooling of condenser on the first law thermal performance(COP) of vapour compression refrigeration using liquid vapour heat exchanger using ecofriendly refrigerants. (For T_eva=253K, T_Cond=323K Compressor efficiency=0.8, Heat exchanger effectiveness=0.80, ,T_super_heating_evaporator=10

Sub cooling Temp	COP_Actual R134a	COP_Actual R1234ze	COP_Actual R1234yf
0	1.922	1.889	1.755
5	2.168	2.141	2.075
10	2.449	2.428	2.370
15	2.774	2.761	2.649
20	3.159	3.154	3.047

Table-4(b) Effect of subcooling of condenser on the exergy Destruction ratio of vapour compression refrigeration using liquid vapour heat exchanger using ecofriendly refrigerants. (For T_eva=253K, T_Cond=323K Compressor efficiency=0.8, Heat exchanger effectiveness=0.80, ,T_super_heating_evaporator=10

Sub cooling Temp	EDR_Rational R134a	EDR_Rational R1234ze	EDR_Rational R1234yf
0	0.6581	0.6640	0.6878
5	0.6143	0.6192	0.6415
10	0.5644	0.5681	0.5891
15	0.5065	0.5089	0.5288
20	0.4382	0.4390	0.4580

Table-4(c) Effect of subcooling of condenser on the second law Efficiency (i.e. exergetic efficiency) thermal performance of vapour compression refrigeration using liquid vapour heat exchanger using ecofriendly refrigerants. (For T_eva=253K, T_Cond=323K Compressor efficiency=0.8, Heat exchanger effectiveness=0.80, ,T_super_heating_evaporator=100.5618

Sub cooling Temp	Exergetic Efficiency R134a	Exergetic Efficiency R1234ze	Exergetic Efficiency R1234yf
0	0.3419	0.3360	0.3122
5	0.3867	0.3808	0.3585
10	0.4356	0.4319	0.4109
15	0.4935	0.4911	0.4712
20	0.5618	0.5610	0.542

Table-5(a) shows the heat exchanger effectiveness variation with first law efficiency (COP) of HFO refrigerants and HFC-134a. The first law efficiency in terms of COP is increasing as heat exchanger effectiveness is increasing. It is clear that COP of HFC-134a is higher than HFO refrigerants. Although COP of R1234ze is similar to R134a but higher than R1234yf. Table-5(b) shows the effectiveness of heat exchanger variation with rational exergy destruction ratio ($EDR_{_Rational}$) of HFO refrigerants and HFC-134a. The effectiveness of heat exchanger is increasing as rational exergy destruction ratio ($EDR_{_Rational}$) is decreasing. It is clear that rational exergy destruction ratio ($EDR_{_Rational}$) of HFC-134a is lower than HFO refrigerants. The high ($EDR_{_Rational}$) is found by using R1234yf as compared to HFC-134a and HFO-1234ze refrigerants. Although rational exergy destruction ratio ($EDR_{_Rational}$) of R1234ze is slightly higher than to R134a. Table-5(c) shows the variation of effectiveness of heat exchanger temperature variation of second law efficiency (exergetic efficiency) HFO refrigerants and HFC-134a. The second law efficiency (exergetic efficiency) is increasing as effectiveness of heat exchanger is increasing. It is clear that second law efficiency (exergetic efficiency) of HFC-134a is higher than HFO refrigerants. Although second law efficiency (exergetic efficiency) of R1234ze is slightly lower to R134a but higher than R1234yf.

Table-5(a) Variation of Effectiveness of heat exchanger with thermal performance of vapour compression refrigeration using liquid liquidvapour heat exchanger using R134a ecofriendly refrigerant. (For T_Eva=253 K ,T_cond=323K, Compressor efficiency=0.8, T_superheat_Eva=10, T_sub_cool_Cond=5

Effectiveness of heat exchanger	$COP_{_Actual_R134a}$	$COP_{_Actual_R1234ze}$	$COP_{Actual_R1234yf}$
0.40	2.390	2.36	2.234
0.45	2.398	2.368	2.243
0.5	2.405	2.377	2.253
0.55	2.412	2.385	2.262
0.6	2.420	2.394	2.272
0.65	2.427	2.402	2.281
0.70	2.434	2.411	2.291
0.75	2.434	2.42	2.30
0.80	2.449	2.428	2.31
0.85	2.456	2.437	2.319
0.90	2.463	2.445	2.329
0.95	2.471	2.454	2.338
1.0	2.478	2.462	2.348

Table-5(b) Variation of Effectiveness of heat exchanger with thermal performance of vapour compression refrigeration using liquid liquidvapour heat exchanger using R1234ze ecofriendly refrigerant. (For T_cond=323K, T_Eva=253 K,Compressor efficiency=0.8, T_superheat_Eva=10, T_sub_cool_Cond=5

Effectiveness of heat exchanger	$EDR_{_Rational\ R134a}$	$EDR_{_Rational\ R1234ze}$	$EDR_{_Rational\ R1234yf}$
0.40	0.5748	0.5803	0.6027
0.45	0.5735	0.5788	0.6010
0.5	0.5722	0.5772	0.5993
0.55	0.5709	0.5757	0.5976
0.6	0.5696	0.5742	0.5959

0.65	0.5683	0.5727	0.5942
0.70	0.5670	0.5712	0.5925
0.75	0.5657	0.5696	0.5908
0.80	0.5644	0.5681	0.5891
0.85	0.5631	0.5666	0.5874
0.90	0.5618	0.5651	0.5858
0.95	0.5605	0.5636	0.5841
1.0	0.5592	0.5620	0.5824

Table-5(c) Variation of Effectiveness of heat exchanger with thermal performance of vapour compression refrigeration using liquid liquidvapour heat exchanger using R1234yf ecofriendly refrigerant. (For T_cond=323K, T_Eva=253 K,Compressor efficiency=0.8, T_superheat_Eva=10, T_sub_cool_Cond=5

Effectiveness of heat exchanger	Exergetic Efficiency	Exergetic Efficiency	Exergetic Efficiency
0.40	0.4252	0.4197	0.3973
0.45	0.425	0.4212	0.3990
0.5	0.4278	0.4229	0.4007
0.55	0.4291	0.4243	0.4024
0.6	0.4304	0.4258	0.4041
0.65	0.4317	0.4273	0.4058
0.70	0.4330	0.4288	0.4075
0.75	0.4343	0.4304	0.4092
0.80	0.4356	0.4319	0.4109
0.85	0.4369	0.4334	0.4126
0.90	0.4382	0.4349	0.4142
0.95	0.4395	0.4364	0.4159
1.0	0.4408	0.4380	0.4176

Table-6(a) shows the compressor efficiency variation with first law efficiency (COP) of HFO refrigerants and HFC-134a. The first law efficiency in terms of COP is increasing as Compressor efficiency is increasing. It is clear that COP of HFC-134a is higher than HFO refrigerants. Although COP of R1234ze is similar to R134a but higher than R1234yf. Table-6(b) shows the Compressor efficiency variation with rational exergy destruction ratio ($EDR_{Rational}$) of HFO refrigerants and HFC-134a. The Compressor efficiency is increasing as rational exergy destruction ratio ($EDR_{Rational}$) is decreasing. It is clear that rational exergy destruction ratio ($EDR_{Rational}$) of HFC-134a is lower than HFO refrigerants. The high ($EDR_{Rational}$) is found by using R1234yf as compared to HFC-134a and HFO-1234ze refrigerants. Although rational exergy destruction ratio ($EDR_{Rational}$) of R1234ze is slightly higher than to R134a. Table-6(c) shows the variation of compressor efficiency temperature variation of second law efficiency (exergetic efficiency) HFO refrigerants and HFC-134a. The second law efficiency (exergetic efficiency) is increasing as compressor efficiency is increasing. It is clear that second law efficiency (exergetic efficiency) of HFC-134a is higher than HFO refrigerants. Although second law efficiency (exergetic efficiency) of R1234ze is slightly lower to R134a but higher than R1234yf.

Table-6(a) Variation of compressor efficiency with First law thermal performance(COP) of vapour compression refrigeration using liquid liquidvapour heat exchanger using ecofriendly refrigerant.s (For T_cond=323K, T_Eva=253 K, Effectiveness of heat exchanger =0.8, T_superheat_Eva=10, T_sub_cool_Cond=5

Compressor efficiency	COP_{Actual} R134a	COP_{Actual} R1234ze	COP_{Actual} R1234yf
0.6	1.626	1.606	1.512
0.65	1.762	1.739	1.638
0.7	1.897	1.873	1.764
0.75	2.039	2.007	1.889
0.80	2.168	2.141	2.015
0.85	2.304	2.275	2.141
0.90	2.439	2.409	2.267
0.95	2.575	2.542	2.393
1.0	2.710	2.675	2.519

Table-6(b) Variation of compressor efficiency with thermal performance in terms of rational exergy destruction ratio of vapour compression refrigeration using liquid vapour heat exchanger using ecofriendly refrigerants. (For T_cond=323K, T_Eva=253 K, Effectiveness of heat exchanger =0.8, T_super heat_Eva=10, T_sub_cool_Cond=5

Compressor efficiency	$EDR_{Rational}$ R134a	$EDR_{Rational}$ R1234ze	$EDR_{Rational}$ R1234yf
0.6	0.7108	0.7144	0.7311
0.65	0.6867	0.6906	0.7087
0.7	0.6626	0.6668	0.6863
0.75	0.6384	0.6430	0.6639
0.80	0.6145	0.6182	0.6415
0.85	0.5902	0.5954	0.6191
0.90	0.5661	0.5716	0.5967
0.95	0.542	0.5478	0.5743
1.0	0.5179	0.524	0.5519

Table-6(c) Variation of compressor efficiency with second thermal performance(exergetic Efficiency) of vapour compression refrigeration using liquid liquidvapour heat exchanger using ecofriendly refrigerants. (For T_cond=323K, T_Eva=253 K, Effectiveness of heat exchanger =0.8, T_superheat_Eva=10, T_sub_cool_Cond=5

Compressor efficiency	Exergetic Efficiency R134a	Exergetic Efficiency R1234ze	Exergetic Efficiency R1234yf
0.6	0.2892	0.2856	0.2689
0.65	0.3133	0.3094	0.2913
0.7	0.3374	0.3332	0.3137
0.75	0.3616	0.357	0.3361
0.80	0.3857	0.3806	0.3585
0.85	0.4098	0.4046	0.3809
0.90	0.4339	0.4284	0.4033
0.95	0.4580	0.4522	0.4257
1.0	0.4821	0.4760	0.4481

4. Conclusion

The following conclusions were drawn from present investigations.

(i) The first law efficiency in terms of COP is decreasing as condenser temperature is increasing. The first law efficiency (COP) of HFC-134a is higher than HFO refrigerants. The first law efficiency in terms of COP of R1234ze is similar to R134a but higher than R1234yf.

(ii) The condenser temperature is increasing as rational exergy destruction ratio ($EDR_{Rational}$) is also increasing. (v)The second law efficiency (exergetic efficiency) is decreasing as condenser temperature is increasing. The exergetic efficiency of HFC-134a is higher than HFO refrigerants. Although exergetic efficiency of R1234ze is slightly less than R134a but higher than R1234yf.

(iii) The first law efficiency in terms of COP is increasing as evaporator temperature is increasing. The evaporator temperature is increasing as rational exergy destruction ratio ($EDR_{Rational}$) is also increasing. The rational exergy destruction ratio ($EDR_{Rational}$) of HFC-134a is lower than HFO refrigerants. The highest ($EDR_{Rational}$) is found by using R1234yf as compared to HFC-134a and HFO-1234ze refrigerants. Rational exergy destruction ratio ($EDR_{Rational}$) of R1234ze is slightly higher than to R134a.

(iv) The second law efficiency (exergetic efficiency) is decreasing as evaporator temperature is increasing. The second law efficiency of HFC-134a is higher than HFO refrigerants. Although second law efficiency of R1234ze is slightly lower than R134a but higher than R1234yf.

(v) The first law efficiency in terms of COP is increasing as super heating temperature is increasing and rational exergy destruction ratio ($EDR_{Rational}$) is decreasing. It is clear that rational exergy destruction ratio ($EDR_{Rational}$) of HFC-134a is lower than HFO refrigerants.

(vi) The second law efficiency (exergetic efficiency) is increasing as super heating temperature is increasing. It is clear that second law efficiency of HFC-134a is higher than HFO refrigerants. Although second law efficiency of R1234ze is similar to R134a but higher than R1234yf.

(vii) The first law efficiency in terms of COP is increasing as subcooling temperature is increasing. The sub cooling temperature is increasing as rational exergy destruction ratio ($EDR_{Rational}$) is also decreasing. The rational exergy destruction ratio ($EDR_{Rational}$) of HFC-134a is lower than HFO refrigerants. The high ($EDR_{Rational}$) is found by using R1234yf as compared to HFC-134a and HFO-1234ze refrigerants.

(viii) The second law efficiency (exergetic efficiency) is increasing as sub cooling temperature is increasing. It is clear that second law efficiency (exergetic efficiency) of HFC-134a is higher than HFO refrigerants. Although second law efficiency (exergetic efficiency) of R1234ze is slightly lower to R134a but higher than R1234yf.

(ix) The first law efficiency in terms of COP is increasing as heat exchanger effectiveness is increasing. It is clear that COP of HFC-134a is higher than HFO refrigerants. Although COP of R1234ze is similar to R134a but higher than R1234yf.

(x) The effectiveness of heat exchanger is increasing as rational exergy destruction ratio ($EDR_{Rational}$) is decreasing. It is clear that rational exergy destruction ratio ($EDR_{Rational}$) of HFC-134a is lower than HFO refrigerants. The high ($EDR_{Rational}$) is found by using R1234yf as compared to HFC-134a and HFO-1234ze refrigerants. Although rational exergy destruction ratio ($EDR_{Rational}$) of R1234ze is slightly higher than to R134a.

(xi) The second law efficiency (exergetic efficiency) is increasing as effectiveness of heat exchanger is increasing. It is clear that second law efficiency (exergetic efficiency) of HFC-134a is higher than HFO refrigerants. Although second law efficiency (exergetic efficiency) of R1234ze is slightly lower to R134a but higher than R1234yf.

(xii) The first law efficiency in terms of COP is increasing as Compressor efficiency is increasing. It is clear that COP of HFC-134a is higher than HFO refrigerants. Although COP of R1234ze is similar to R134a but higher than R1234yf.

(xiii) The second law efficiency (exergetic efficiency) is increasing as compressor efficiency is increasing. It is clear that second law efficiency (exergetic efficiency) of HFC-134a is higher than HFO refrigerants.

(xiv) Second law efficiency (exergetic efficiency) of R1234ze is slightly lower to R134a but higher than R1234yf. The compressor efficiency is increasing as rational exergy destruction ratio ($EDR_{Rational}$) is decreasing.

(xv) The rational exergy destruction ratio ($EDR_{Rational}$) of HFC-134a is lower than HFO refrigerants. The high ($EDR_{Rational}$) is found by using R1234yf as compared to HFC-134a and HFO-1234ze refrigerants. Although rational exergy destruction ratio ($EDR_{Rational}$) of R1234ze is slightly higher than to R134a.

References

[1] Johnson. E. Global warming from HFC. Environ. Impact Asses.1998; 18:485-492.

[2] Esbri N., Miranda J. M. M., Babiloni A.M., Cervera A. B., Flores J. M. B., "Experimental analysis of R1234yf as a drop in replacement for R134a in a vapour compression system", International Journal of Refrigeration 30 (2013) Pp. 01-11.

[3] Zhao Y., Liang Y., SunY., Chen J., "Development of a mini-channel evaporator model using R1234f as working fluid., International Journal of Refrigeration 35 (2012) Pp. 2166-2178.

[4] C. Nikolaidis, D. Probert-Exergy method analysis of a two-stage vapour-compression refrigeration-plants performance. Int J Applied Thermal Engineering.1998; 60:241-256

[5] H.M Getu, P.K Bansal. Thermodynamic analysis of an R744-R717 cascade refrigeration system. Int J Refrigeration.2008; 31:45-54.

[6] Mark W. Spatz, Samuel F. Yana Motta. An evaluation of options for replacing HCFC-22 in medium temperature refrigeration systems. Int J Refrigeration.2004; 27:475-483.

[7] M. Mohanraj, et.al. Experimental investigation of R290/R600a mixture as an alternative to R134a in a domestic refrigerator. Int J Thermal Sciences.2009; 48:1036-1042.
[8] Mahmood Mastani Joybari,et.at, Exergy analysis and optimization of R600a as a replacement of R134a in a domestic refrigerator system. International Journal of refrigeration.2013; 36:1233-1242
[9] X.H. Han. et.al, Cycle performance study on R32/R125/R161 as an alternative refrigerant to R407C. Int J Applied Thermal Engineering.2007; 27:2559-2565.
[10] E. Halimic, D. Ross, B. Agnew, A. Anderson, I. Potts. A comparison of the operating performance of alternative refrigerants. Int J Applied Thermal Engineering.2003; 23:1441-1451.
[11] J.U. Ahamed, R.Saidur, H.H.Masjuki. A review on exergy analysis of vapor compression refrigeration system. Int J Renewable and sustainable energy reviews.2011; 15:1593-1600.
[12] J.U. Ahamed, R.Saidur, H.H.Masjuki. A review on exergy analysis of vapor compression refrigeration system. Int J Renewable and sustainable energy reviews.2011; 15:1593-1600.
[13] J.U Ahamed, R.saidur, H.H Masjuki, M.A Sattar- An analysis of energy, exergy and sustainable development of a vapour compression refrigeration system using hydrocarbon.International journal of Green energy.2012;9:707-717.
[14] V. Siva Reddy, N.L Panwar, S.C Kaushik-Exergy analysis of a vapour compression refrigeration system withR134a, R143a, R152a, R404A, R407C, R410A, R502 and R507A.Clean Techn Environ Policy.2012;14:47-53.
[15] R.Saravanakumar, V.Selladurai-Exergy analysis of a domestic refrigerator using eco-friendly R290/R600a refrigerant mixture as an alternative to R134a.Int J Therm Anal Calorim.2013;
[16] S.Kumar, M.Prevost, R.Bugarel- Exergy analysis of a vapour compression refrigeration system. Heat Recovery Systems & CHP.1989;9:151-157.
[17] S.Anand ,S.KTyagi-Exergy analysis and experimental study of a vapour compression refrigeration cycle.Int J Therm Anal Calorim.2012; 110:961-971.
[18] X.H. Han, Q. Wang, Z.W. Zhu, G.M. Chen. Cycle performance study on R32/R125/R161 as an alternative refrigerant to R407C. Int J Applied Thermal Engineering.2007; 27:2559-2565.
[19] R. Cabello, J. Navarro-Esbrı, R. Llopis, E. Torrella. Analysis of the variation mechanism in the main energetic parameters in a single-stage vapour compression plant. Int J Applied Thermal Engineering.2007; 27:167-176.
[20] R.Cabello a, E.Torrella b, J.Navarro-Esbr. Experimental evaluation of a vapour compression plant performance using R134a, R407C and R22 as working fluids. Int J Applied Thermal Engineering.2004; 24:1905-1917.
[21] Camelia Stanciu, Adina Gheorghian, Dorin Stanciu, Alexandru Dobrovicescu- Exergy analysis and refrigerant effect on the operation and performance lomits of a one stage vapour compression refrigeration system.Termotehnica.2011;1:36-42
[22] B.O. Bolaji, M.A. Akintunde, T.O. Falade. Comparative analysis of performance of three ozone-friends HFC refrigerants in a vapor compression refrigerator. Int J Sustainable Energy & Environment.2011; 2:61-64.
[23] RecepYumrutas, Mehmet Kunduz, Mehmet Kanoglu-Exergy analysis of vapor compression refrigeration systems. Exergy, An International Journal.2002; 2:266-272.
[24] M. Padilla, R. Revellin, J. Bonjour. Exergy analysis of R134a as replacement of R12 in a domestic refrigeration system. Int J Energy Conversion and Management.2010; 51:2195-2201.
[25] Akhilesh Arora, S.C. Kaushik. Theoretical analysis of a vapour compression refrigeration system with R502, R404A and R507A. Int J Refrigeration.2008; 31:998-1005

Cite this article as: R.S. Mishra, Performance improvement of vapour compression refrigeration systems, International journal of research in engineering and innovation (IJREI), vol 3, issue 1 (2019), 48-56.

Natural Hazards
https://doi.org/10.1007/s11069-018-03568-y

ORIGINAL PAPER

Physical vulnerability assessment of buildings exposed to landslides in India

Aditi Singh[1] · D. P. Kanungo[2] · Shilpa Pal[3]

Received: 8 August 2018 / Accepted: 31 December 2018
© Springer Nature B.V. 2019

Abstract

Safe structures are the backbone of human coping capacity towards healthy living that can contribute significantly in reducing risk during hazards. However, due to various natural and anthropogenic activities, about 12.6% of land areas (excluding snow-covered area) in India are prone to landslide posing threat to life and property. Moreover, many structures in the hilly terrain of India are non-engineered which results in high vulnerability of buildings. Therefore, assessment of physical vulnerability is a fundamental step in reducing landslide risk. The study aims to present a methodology to assess vulnerability of the buildings using indicator-based approach at site-specific scale. Several studies to assess vulnerability of buildings due to landslides have been carried out by researchers from European countries. But these methodologies cannot be implemented successfully in India because of different geological and climatic condition. The different components of the discussed methodology for physical vulnerability of buildings exposed to landslides such as landslide intensity (a function of landslide velocity and volume) and resistance of buildings (a function of structural and non-structural features) are worked out and suggested by different researchers. However, putting them together, to present as a framework (specifically in Indian scenario) is the novelty of the present work. Further, consideration of the concept of 'proximity of buildings to landslides' in the process of site-specific vulnerability assessment is newly proposed. To address this issue, fifteen potential indicators contributing to vulnerability of buildings have been identified and a systematic form for documentation of data during field survey has also been prepared (keeping in view the construction bye-laws and techniques followed in India). The methodology discussed is further successfully implemented in ward number 10 of Gopeshwar Township (Chamoli District), Uttarakhand, India.

Keywords Landslide · Physical Vulnerability · Site-Specific Vulnerability · Indicators · Buildings

✉ Aditi Singh
aditi.peace@gmail.com

[1] Civil Engineering Department, Gautam Buddha University, Greater Noida, India

[2] Geotechnical Engineering Group, CSIR—Central Building Research Institute (CBRI), Roorkee, Uttarakhand, India

[3] Department of Civil Engineering, Delhi Technological University, Delhi, India

Published online: 14 January 2019

1 Introduction

Landslides form one of the most damaging and life-threatening hazards, which affect most of the hilly regions worldwide. According to the Global Assessment Report on Disaster Risk Reduction (ISDR 2009, United Nations, Geneva, Switzerland), the total number of people potentially exposed to landslides is estimated to be greatest in India, followed by Indonesia, China, Philippines, Japan and Taiwan Province of China. In India, about 0.42 million sq. km or 12.6% of land area, excluding snow-covered area, is prone to landslide hazard (www.gsi.gov.in, retrieved on 07th August, 2018). The Himalayas and the Western Ghats are the two most vulnerable regions to landslide disasters in India.

Safe structures are the backbone of human copying capacity towards healthy living that can contribute significantly in reducing risk during hazards such as landslides. However, in the hilly terrain of India, most of the buildings are non-engineered and are constructed on the hill slope with irregular structural configuration. The vulnerability of buildings is magnified due to exponential growth in population and lenient construction bye-laws. The lacuna in construction quality in landslide-prone region also exists since there are no detailed bye-laws specific to landslides as in case of earthquake (i.e. IS: 4326-2013, for earthquake resistant design and construction of buildings). Further, the extremely changeable characteristics of landslides such as volume, velocity and magnitude and lack of historical data in context to such hazards in most of the cases (specifically in Indian scenario), makes the assessment of vulnerability of buildings a complex task.

Therefore, considering the ever-increasing incidences of landslide induced tragedies in the Himalaya, a methodology has been proposed for Indian scenarios to assess the vulnerability of buildings at site-specific scale using indicator-based approach. In this study, the indicators are selected keeping in view the Indian hilly terrain. Then, the physical vulnerability (PV) of buildings in the study area is quantified as function of intensity of landslide and resistance of buildings. The resistance of buildings will ensure the integrity and functionality of buildings at different intensity of landslides. The database of buildings in the study area is prepared by field inspection. Further, for recording field data more systematically during field survey, a documentation form for Building Vulnerability Assessment (BVA) is prepared. Based on the available literature and the codal provisions and guidelines followed in India, weights are assigned to each indicator accordingly. The site-specific vulnerability (SSV) of buildings is quantified using some well-established empirical relationships from the literature. The proposed model is implemented by taking an actual landslide (debris flow), surrounded by dense built-up area located in ward number 10 in Gopeshwar township of Chamoli district, Uttarakhand (India).

The methodology discussed and implemented here is suitable in those areas where the availability of historical data on landslide (such as velocity, volume, intensity) is scarce and where relevant information on the types of buildings or the vulnerability curves on different types of buildings does not exist. Such type of studies can be extensively used by engineers in preparing the construction guidelines specific to a particular region and also by scientists to facilitate suitable protective mitigation measures.

2 Review of existing methods of vulnerability assessment of buildings for landslides

The term vulnerability is used by scientists from various disciplines in a number of ways. From a natural and engineering science perspective, vulnerability may be defined as the degree of loss, in terms of percentage of structural damage, to a given element or set of elements subjected to a landslide event of a given type and intensity (Glade 2003). Nevertheless, vulnerability curves (Fuchs et al. 2007; Totschnig et al. 2011), vulnerability matrices (Papathoma-Kohle et al. 2012) and vulnerability indicators (Birkmann et al. 2013) are the three most commonly used methods for assessing physical vulnerability. These approaches are based on general assumptions and each approach manifests some pros and shortcomings. Vulnerability matrices are qualitative methods, and the development of matrices is often dependent on expert judgement or empirical data (Papathoma-Köhle et al. 2017). According to Menoni (2006), vulnerability matrices are composed by classified intensities and stepwise damage levels. While vulnerability curves are usually building specific and link the intensity of the landslide hazard directly, providing quantitative results (Papathoma-Köhle 2016). Moreover, both the methods neglect the characteristics of the buildings that contribute to their vulnerability. A brief review on these two approaches can be found in the works of Singh et al. (2017) and Papathoma-Köhle et al. (2017). However, the third method 'vulnerability indicator' is in its infancy and a limited research has been carried out in context of landslides to assess vulnerability of buildings with indicators (Papathoma-Köhle et al. 2007; Kappes et al. 2012; Silva and Pereira 2014; Thouret et al. 2014; Ettinger et al. 2016). Since lot of uncertainties is associated in using indicator-based approach, some studies use them in combination with well-established approach such as with vulnerability curves (Rheinberger et al. 2013; Du et al. 2014; Godfrey et al. 2015).

In order to know the likely impact of a landslide prior to the event on buildings and infrastructure in any part of a country, it is essential to assess the vulnerability of the built environment of the affected regions. Thus, indicator-based approach serves as a useful method to assess the structural integrity of a building by focusing on the individual indicator affecting the building during landslide. This method will provide an insight to planners, researchers and engineers on the indicators that make structures vulnerable during landslide hazard and will help them take necessary actions beforehand by retrofitting the existing structures and providing guidelines for future construction works.

2.1 Indicator-based approach for vulnerability assessment of buildings exposed to landslides—an international scenario

Assessing vulnerability of buildings in a landslide-prone region is a primary component in reducing landslide risk. Scientists and researchers in the field of landslide are more focussed on research related to landslide susceptibility and hazard and limited studies have been conducted on assessing vulnerability of buildings due to landslides. The construction practices also vary in different parts of the country even when using the same construction material. There is no standard method yet to determine vulnerability of buildings using indicator-based approach. The detailed review on assessing the vulnerability of buildings using indicators is discussed in this section.

As far as mountain hazards are concerned, Papathoma-Köhle et al. (2007) made one of the first attempts to relate physical vulnerability of buildings to landslides. Five indicators

(material, surrounding, slope side, warning and floors) were used to assess vulnerability of each building based on expert judgement by assigning weights to each indicator according to their importance.

Uzielli et al. (2008) and Li et al. (2010) presented a quantitative model for assessing vulnerability. Uzielli et al. (2008) assessed vulnerability in terms of landslide intensity and the susceptibility of vulnerable elements such as structures and persons; however, Li et al. (2010) assessed vulnerability based on landslide intensity and resistance of exposed elements (i.e. buildings).

Chen et al. (2011) discussed about the semi-quantitative evaluation of building vulnerability using Zhaoshuling landslide located in the Three Gorges Reservoir Area, China. The study carried out detailed field investigation on the geological condition of landslide after detailed field investigation on the buildings location and structure.

Kappes et al. (2012) presented a study on vulnerability assessment exposed to multi-hazards (avalanche, rockfall, flood, shallow landslides, debris flow and flash floods) using indicator-based method in alpine environments. Vulnerability indicators (based on building-specific information and building surroundings) such as materials, floors, condition, opening towards slope (size and condition), height of lowest opening, presence of warning signs of landslides, basement, roof material and foundation type, building row (towards slope and towards river, protection by vegetation, protection measures, movable objects that can be carried away by water or snow) were provided in this study.

Puissant et al. (2013) proposed a GIS-aided indicator-based methodology with an application to regional-scale consequence analysis. This method allows estimation of the possible damage caused by landslides by combining weighted indicators reflecting the exposure of the elements at risk. The method has been successfully applied to analyse the present consequences in the Barcelonnette Basin (South French Alps).

Silva and Pereira (2014) assessed the physical vulnerability of buildings to shallow landslides by using indicators such as construction technique and material, floor and roof structure, number of floors etc. to calculate the resistance of the buildings.

Uzielli et al. (2015) presented a study on the estimation of risk around the Ancona landslide for 39 buildings. In this study quantitative estimation of a set of buildings was based on the intensity of landslide and resilience of the building. Some common indicators such as foundation type, age and structural type was used as a function of resilience.

Godfrey et al. (2015) developed an expert based approach that allowed assessment of physical vulnerability of buildings to hydro-meteorological hazards in areas having limited information about the hazard or the exposed elements. Indicators like floor height from ground level, number of floors, building size, foundation type, foundation depth, building maintenance, presence of basement, cracks in the structure, wall around the building and building located on slope and etc. were used in case of debris flow to calculate the vulnerability index. The weighing process involved pre-selection of 20 respondents having prior knowledge in field of natural hazard and risk assessment.

Papathoma-Köhle (2016) presented comparison of two approaches, i.e. vulnerability curves and indicator-based methodology for debris flow hazards. The study emphasised on the need to improve the existing methodology and to combine the two approaches to exploit their full power. Further, an indicator-based methodology (IBM) for mountain hazards including debris flow (Kappes et al. 2012) is applied to a case study for debris flows in South Tyrol, where in the past a vulnerability curve has been developed.

Table 1 presents a list of indicators which has been used by various researchers and scientists in the past to assess vulnerability of buildings due to landslides. It is interesting to note that most of the researches on assessing physical vulnerability are carried out by

Table 1 Summary of indicators used in the literature to assess vulnerability of buildings

	Indicators	Papathoma-Köhle et al. (2007)	Uzielli et al. (2008)	Li et al. (2010)	Chen et al. (2011)	Kappes et al. (2012)	Puissant et al. (2013)	Silva and Pereira (2014)	Uzielli et al. (2015)	Godfrey et al. (2015)	Papathoma-Kohle (2016)	Total no. of literature
		L	L	L	L	MH	L	L	L	HMH	L	
1	Construction material	X				X	X	X			X	5
2	Number of floors	X		X	X	X	X	X		X	X	8
3	Surrounding wall	X								X	X	3
4	Slope side									X		1
5	Presence of large windows on slope side	X				X						2
6	Presence of basement					X				X	X	3
7	Foundation type			X					X	X		3
8	Foundation depth					X				X		2
9	Presence of warning sign	X				X					X	3
10	State of maintenance/conservation status		X	X		X	X	X		X		6
11	Structural typology/technique		X	X	X			X	X	X		6
12	Construction Year/age				X		X		X			3
13	Building Function				X		X					2

Table 1 (continued)

	Indicators	Papathoma-Köhle et al. (2007)	Uzielli et al. (2008)	Li et al. (2010)	Chen et al. (2011)	Kappes et al. (2012)	Puissant et al. (2013)	Silva and Pereira (2014)	Uzielli et al. (2015)	Godfrey et al. (2015)	Papathoma-Kohle (2016)	Total no. of literature
		L	L	L	L	MH	L	L	L	HMH	L	
14	Roof structure/roof material					X		X				2
15	Floor height					X				X	X	3
16	Number of openings									X	X	2
17	Building size									X		1
18	Wall material									X		1
19	Quality of Construction									X		1
20	Cracks in structure									X		1
21	Building location—(on a slope; close to slope; etc.)					X				X	X	3
22	Presence of in between buildings									X		1
23	Protection by vegetation					X					X	2

X: Indicators used by various researchers in the past

Hazard type:*L* landslide, *MH* Multi-hazard, *HMH* hydro-meteorological hazards

European countries and the indicators selected in their studies are based on the construction techniques followed in their country and also on the local factors of the affected area (which may or may not be same in other landslide-prone regions). Some of the indicators like number of floors, structural typology/technique, and state of maintenance/conservation status were common in most of the studies; however, only few used maximum number of indicators in their study.

2.2 Physical vulnerability assessment due to landslides in India

Most regions of hilly terrains in India fall under seismic zones IV and V and are susceptible to natural calamities (such as heavy rainfall, high speed winds, cloud bursts, landslides and flash floods). Amongst all these natural disasters, landslides is one such disaster occurring frequently in Indian Himalayas, which lead to significant loss of life and property annually (Sarkar et al. 2015). Due to rapid growth in population and influx of tourist in the hilly areas, tremendous development has taken place from the last few decades in the hill towns of India. Majority of buildings on hill slopes have unequal column heights, i.e. the building configuration is irregular with foundations at different levels. Such type of construction makes the buildings more vulnerable to landslide hazard. Further, lenient construction bye-laws and low awareness towards structural safety amongst general public has made the buildings more fragile.

It is recognised that most of the studies carried out by various researchers and scientists of governmental and non-governmental organisations in India are limited to hazard and susceptibility zonation (Kanungo et al. 2006; Mathew et al. 2009; Ghosh et al. 2011; Pham et al. 2017 etc.). Very limited detailed information currently exists regarding building typologies in India and the degree of damage of buildings in the affected areas. Additionally, efforts and data collection are focused mainly on the human casualties due to the event itself and not the damaged buildings, which might also play a significant role (Douglas 2007). Therefore, in order to understand the behaviour of a building (i.e. the likely impact and degree of damage at varying intensities) in post-landslide event, a formal method is required to comprehensively assess resistance ability of individual houses on the basis of their structural and geographical characteristics in the affected areas.

Thennavan et al. (2016) presented a study to assess the vulnerability of buildings under landslide hazard-prone slopes of the Nilgiris district of India by utilising the methodology of Papathoma-Köhle et al. (2007). However, the type of landslide (whether shallow landslide or debris flows) is not clearly discussed in this study.

Singh et al. (2018) proposed a semi-quantitative approach for estimation of physical vulnerability of buildings exposed to different landslide intensity scenarios. Physical vulnerability of buildings is estimated based on the landslide intensity and resistance of the buildings using five indicators and further classified as class I, II or III for scenario-based low to very high landslide intensity. The methodology has been illustrated through a case study involving 71 buildings located in the Joshimath Block of the Chamoli District in Uttarakhand State, India.

Currently, the existing methods to assess vulnerabilities towards distinct processes vary widely and yet no standard method is applicable for landslide hazards (Papathoma-Köhle et al. 2011; Walker and Deeming 2006). Thus, a method is proposed to assess the structural integrity of buildings for Indian scenarios that can be implemented on the ground level. The section below discusses in detail the newly developed methodology for assessing vulnerability of buildings to landslide for Indian scenarios.

3 Methodology for assessing site-specific vulnerability of buildings exposed to landslides

Vulnerability definition may be easy to understand; however, in reality its assessment is complex. Vulnerability assessment depends on several factors such as (a) damage data from past landslide specifically highlighting details about building characteristics (or indicators), (b) type of landslide and (c) the intensity of landslide which is further associated with the volume and velocity. Such types of data are not always available, which impedes in the validation of any vulnerability model. Therefore, vulnerability of buildings cannot be assessed adequately and exclusively due to lack of data on the past events (Birkmann 2006). Hence, a simple and user-friendly methodology that explains the identification and selection of indicators for assessing physical vulnerability of buildings to preserve its integrity and functionality amid varying intensity of landslide is important. In order to assess the condition of building across the hilly terrain, Building Vulnerability Assessment (BVA) to landslides becomes an essential tool. BVA should be undertaken to identify the buildings and vital infrastructures which require special attention in order to make them more resistant against the landslide disaster. The proposed methodology is designed to assess individual building based on a number of indicators that needs to be taken into consideration for carrying out physical vulnerability assessment of buildings for Indian scenarios. The inspection of building and documentation of data in the form may require 15 to 30 min, depending upon the size of the building. BVA is a five stage process, which includes:

(i) Selection of the study area and acquisition of hazard information
(ii) Selection criteria for indicators
(iii) Preparation of documentation form
(iv) Normalisation and standardisation of indicator scores
(v) Quantification of vulnerability at site-specific scale

3.1 Selection of the study area and acquisition of hazard information

Increase in population in mountainous regions has resulted in developmental activities even on the steep slopes. For this reason, the buildings and structures constructed in landslide-prone areas should be considered for assessing their vulnerability due to landslides. Therefore, the selection of the study area is based entirely on the objective of the work and the scale and extent of the study area (village, township, municipalities, regions, or countries), since the choice of the mapping scale affects the selection of the approach (Aleotti and Chowdhury 1999). Remote sensing data, field data and Geographical Information System (GIS) can be used to acquire and analyse information on landslide hazard of that particular region. Detailed information regarding landslide can be obtained by investigation of historical records, field surveys and aerial photo-interpretation and then developing a landslide inventory of the study area. Nevertheless, it is often advised never to rely entirely on a single source but to combine as many of the acquisition techniques as possible to obtain accurate data.

In the present study, Gopeshwar township (the headquarter of Chamoli District in Uttarakhand State, India) is selected for physical vulnerability assessment of buildings as the hill township encounters landslide disasters in different parts. To achieve this objective, the landslide hazard information for the study area is collected and mapped from the available landslide inventory map at macro-scale for the upper Alaknanda valley (Sajwan and Sushil 2016), Google Earth Image and extensive field surveys using GIS tools. For the site-specific vulnerability assessment, a possible debris flow in Ward No. 10 of Gopeshwar township is selected as a case study for implementation of the methodology.

3.2 Selection criteria for indicators

Many structural and geographical characteristics of buildings play an important role in influencing their vulnerability during a landslide disaster. From the literature, it can be found that vulnerability of a building to landslides depends on its resistance which is subject to its building typology, age and height (Dai et al. 2002). Totschnig et al. (2011) referred to indicators such as number of storeys, construction materials, existence of a basement and local protection measures such as reinforced outer walls and sheltered openings as key indicators for assessing physical vulnerability of buildings specifically in case of a debris flow. Therefore, building indicators for assessing vulnerability needs to be carefully selected keeping in view the hilly terrain of Himalayas. The selected indicators to assess physical vulnerability of buildings exposed to landslides are based on the literature review and the Indian Standard (IS: 14804-2000) on 'Siting, Design and Selection of Materials for Residential Buildings in Hilly Areas—Guidelines' and National Building Code of India (NBC: 2016, Volume-1, Annexure F, Clause 16) on 'Special requirements for development planning in hilly areas'. The indicators and the reason of their selection in this study are discussed in Table 2.

3.3 Preparation of documentation form

After the identification of indicators contributing to the vulnerability assessment of buildings exposed to landslides, a recording form is prepared based on the literature (Cooper 2008; Papathoma-Köhle et al. 2012; Papathoma-Köhle et al. 2015; Del Soldato et al. 2017). Keeping in view the damage scheme and the indicators adopted in the literature, a documentation form is prepared for Indian scenarios by utilising the Google Earth and field surveys information. Details on topographic location are obtained from Google Earth and field surveys helped in obtaining detailed documentation of the buildings along with photographic documentation and interviews with local people and building owners.

The proposed Building Vulnerability Assessment (BVA) form (Table 3) is prepared for systematically collecting the data of buildings during field survey. The form describes the detailed information of the following structural and non-structural indicators during field survey:

(a) *General information* Name, Address and Geographical location (coordinates) of building
(b) *Building Function* Residential, institutional (schools, college, hospital, old age homes, training centre etc.), commercial (offices, guest house, hotels, shops and industrial building, fire station etc.) and mixed (buildings used for multiple purposes such as residential and commercial both)
(c) *Vulnerability Indicators*

- Architectural Features: Shape of the building, dimension of building
- Material Characteristics: Material of wall, floor and roof

- Structural Features: Irregularity in building configuration (horizontal & vertical), bulging of walls and columns, number of floors, presence of cracks, cracks width and their shape, water seepage etc.
- Geographical features: distance of built-up area from steep slope
- Workmanship: quality of construction & state of maintenance.

Vulnerability factors like general information and building function can be obtained from the available records in the district administrative office. However, the information on vulnerability indicators (as listed in Table 2) must be collected for each building on the field as it will play a significant role in determining the overall vulnerability of elements at risk (i.e. buildings) accurately. All the structural and non-structural indicators are listed in the form (Table 3) point-wise, with an aim not to miss any detail during field inspection. The form is developed in a way that it is easy to use in the field and will provide quick vulnerability assessment after compilation.

3.4 Normalisation and standardisation of indicator scores

The indicators selected in this study exhibit certain degree of building resistance that prevents damage or collapse of building during landslide. Since resistance factor assigned to each indicator may be difficult to conceive for lay-person; hence, raw scores are assigned to every indicator. The raw scores are the ranks assigned to each and every single indicator depending upon its contribution in making a building vulnerable during landslide. For example, stone masonry is the most vulnerable during landslides amongst all other typologies; hence, a raw score of 1 is assigned. However, RC buildings with infill walls will contribute less; hence, assigned a raw score of 4. The information required for ranking of the indicators is collected from the available literature, expert judgement, interviews with local people and the codal guidelines followed in the country. Table 4 presents the raw scores assigned to each indicator along with the supporting literature. The second step is to assign resistance factors to all the indicators on the scale of 0 (no resistance) and 1 (maximum resistance). In order to remove uncertainties by assigning the resistance factors based on expert judgment and available literature, standardisation method (Voogd 1983) is used to assign resistance factors. The standardised score thus obtained as per Eq. (1) will give the resistance factor for each indicator.

$$\text{Standardized score} = \text{raw score/maximum raw score} \tag{1}$$

3.5 Parameters for the quantification of vulnerability of buildings

The various parameters considered for quantification of site-specific vulnerability of buildings in the study area are discussed as follows:

3.5.1 Landslide intensity (*I*)

In case of Indian Himalayas, landslide intensity has not been reported so far (Sarkar et al. 2015). In this study, landslide intensity *(I)* is estimated as a function of landslide volume (V_{OL}) and expected landslide velocity (V_{EL}) (Cardinali et al. 2002), as shown in Eq. 2.

$$I = f\left(V_{OL}, V_{EL}\right) \tag{2}$$

Table 2 Vulnerability indicators selected for assessing vulnerability of buildings exposed to landslide hazard in India

Type of vulnerability factors	Vulnerability indicators	Indicator ID	Reasons for selection	Supporting literatures	Considerations in the study
Structural features	Year of construction	R1	Construction technique evolved with the advancement in science and technology Buildings constructed before 2000 are not designed according to codal provision	Papathoma-Köhle et al. (2007), Puissant et al. (2013), Uzielli et al. (2015)	Old construction is assumed to be more vulnerable to landslide in comparison to new construction, since old buildings have completed their designed service life
	Roof design	R2	Prevent the material from entering into the building from top Roof slope prevents material from getting accumulated at top of building	Silva and Pereira (2014), Papathoma-Köhle et al. (2007), IS 14804-(2000) (Clause: 6.4.1 and 6.4.2)	Categorised as (a) Flat roof- wood/bamboo, (b) Pitched roof (slate/local material), (c) Flat/Pitched or Gable roof in GI or Asbestos sheet, (d) Flat roof in Reinforced Concrete (RC) and (e) Pitched or Gable roof
	Foundation (type & depth) and number of floors	R5	Foundation depth reflects the capability to resist displacement Deeper foundation on in situ competent strata offers higher resistance	Ragozin and Tikhvinsky (2000), Li et al. (2010), Uzielli et al. (2015), Godfrey et al. (2015)	Number of floors used as a proxy indicator for foundation type and depth due to unavailability of the data. Categorised as single storey, low-rise, medium-rise and high-rise building
	Vertical configuration of building	R6	Important vulnerability indictor for buildings lying in seismic zones Buildings with simple regular geometry and uniformly distributed mass and stiffness in elevation/plan will suffer much less damage, than buildings with irregular configuration	IS 1893- part 1(2016), IS 14804: (2000) (Clause 6.3.3)	Categorised as regular and irregular
	Horizontal configuration of building	R7	Majority of buildings on hill slopes have unequal column heights	IS 14804: (2000) (Clause 6.3.2); IS 4326: (2013) (Clause 4.4.3)	

Table 2 (continued)

Type of vulnerability factors	Vulnerability indicators	Indicator ID	Reasons for selection	Supporting literatures	Considerations in the study
	Distance between windows & doors from each other	R8	For safer construction practice in hilly terrains	IS 14804: (2000) (Clause 6.1.6 (h))	All doors and windows shall not be less than 1 m from each other
	Type of construction	R14	To ensure safety of occupants and other neighbouring buildings	IS 14804: (2000) (Clause 3.1, 3.2 & 3.3)	Categorised as: (a) Engineered construction, (b) marginally engineered construction and (c) non- engineered construction
	Retrofitted building	R15	Retrofitting is important in case of buildings that have experienced hazards like earthquake in the past.	–	Older buildings that have been repaired and retrofitted are assumed to perform better in case of landslide hazard.
Material characteristics	Structural typology	R3	Different structural typology will experience different levels of damage depending on the intensity of landslide It varies from country to country due to variation in geology, altitude, climatic conditions and economic conditions even when using the same construction material	EPFL (2002), Uzielli et al. (2008), Li et al. (2010), Chen et al. (2011), Silva and Pereira (2014), Uzielli et al. (2015), Godfrey et al. (2015)	Typologies are considered as (a) stone masonry (SM), (b) brick masonry (BM) unreinforced, (c) reinforced concrete (RC) frame with infill walls and (d) mixed construction (buildings with different structural typology)
Architectural characteristics	Surrounding wall of building	R4	Provides protection to the building from the intrusion of the landslide material.	Papathoma-Köhle et al. (2007), Godfrey et al. (2015)	Categorised as (a) no/low surrounding wall, (b) medium surrounding wall and (c) strong high wall

Table 2 (continued)

Type of vulnerability factors	Vulnerability indicators	Indicator ID	Reasons for selection	Supporting literatures	Considerations in the study
Geographical features	Distance between two houses from each other	R9	For safer construction practice in hilly terrains	IS 14804: (2000) (Clause 6.2.4 (a))	No house shall be located closer than 1 m from one another
	Building position from steep slope	R10	Role of topographic location of the building should not be ignored	IS 14804: (2000) (Clause 6.2.4 (b))	No house shall preferably be located closer than 10 m to a steep slope. Therefore, categorised as (a) Greater than 10 m, (b) Less than 10 m and (c) On the steep slope
	Surrounding vegetation	R11	Vegetation has direct impact on the stability of hill slopes Deep-rooted trees and shrubs can reduce the occurrence of shallow rapidly moving landslides and also indirectly obstructing smaller slides and rock fall	Hamilton (2008), Kappes et al. (2012)	Categorised as (a) Barren, (b) Cultivated/Sparsely vegetated and (c) Forest
Workmanship	Quality of construction	R12	Related to the safety of the occupants It is important to ensure the design and construction of a building are according to building codes especially in hazard-prone areas for assessing their vulnerability	Godfrey et al. (2015)	Categorised as (a) very low, (b) low, (c) medium and (d) high
	State of maintenance	R13	Related to the safety of the occupants. Maintenance of a building is necessary to ensure that all the elements of a building are in good state to perform its function	Li et al. (2010), Kappes et al. (2012), Puissant et al. (2013), Silva and Pereira (2014), Uzielli et al. (2015), Godfrey et al. (2015)	

Natural Hazards

Table 3 Building Vulnerability Assessment (BVA) Form

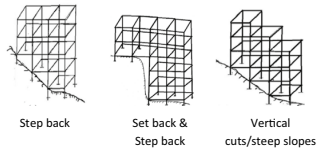
BUILDING LOCATION (COORDINATES)		
ADDRESS		
BUILDING ID		
BUILDING FUNCTION		
S.NO.	**INDICATORS**	
1	Year of Construction, R1	
2	Roof Design, R2	Flat roof in Reinforced Concrete Flat roof in BM/SM (locally available) Flat roof in GI or Asbestos sheet Flat roof in wood/bamboo Pitched or Gable roof
3	Structural Typology, R3	• Stone Masonry • Unreinforced Brick Masonry • RC frames with infill walls • Mixed
4	Surrounding wall of building, R4	• Strong high wall • Medium wall • No/ low surrounding wall
5	Number of Floors, R5	
6	Vertical Configuration (Regular /Irregular), R6	
7	Horizontal Configuration, R7  Step back Set back & Step back Vertical cuts/steep slopes	**1. Regular** (Rectangular shape; L/T/E/Y/ any other shape separated into rectangular parts) **2. Irregular** (step back, set back and step back building, vertical cuts/steep slopes)
8	Presence of windows & doors ≥ 1m from each other (Yes /No), R8	
9	Distance between two houses ≥ 1m from each other (Yes /No), R9	
10	Building Position from Steep slope, R10	Less than 10 m Greater than 10 m On the Steep Slope
11	Surrounding Vegetation, R11	Barren Cultivated Forest
12	Quality of Construction, R12:	

Table 3 (continued)

	(a) Bulging of walls/ column	
	(b) Corrosion in beam/column/slab	
	(c) Spalling of concrete	
	(d) Exposed reinforcement	
	(e) Honeycombing in walls	
	(f) Rebound Hammer test results	
13	State of Maintenance, R13:	
	(A) Cracks in the Structure	
	• due to movement of slope	
	• cracks in exterior member of beams/columns	
	• cracks in plaster material of walls	
	(B) Ground Settlement	
	(C) Water Seepage in Building (Yes /No)	
14	Type of construction, R14	
15	Retrofitted building (Yes /No), R15	
OTHER INFORMATION:		

Since most of the time there is lack of historical data on the intensity parameter, a rough approximation of volume and velocity estimation is discussed in Sects. 3.5.1.1 and 3.5.1.2. Further, after the estimation of volume and velocity, landslide intensity (I) scheme (classified into four classes: low ($I = 0.2$), moderate ($I = 0.4$), high ($I = 0.8$) and very high ($I = 1$), which is proposed in the previous work of author Singh et al. (2018) has been taken into consideration.

3.5.1.1 Landslide volume (V_{OL}) Landslide volume can be estimated using several parameters such as height and width of scar, affected area, shape of landslide through field observations, aerial photographs, satellite images or past reports. However, collection of this information is considered both expensive and challenging. A common way to estimate debris flow volume is by adopting empirical relationships. Table 5 presents a list of empirical relationships used for calculating the volume of landslide. The volume of landslide in this study area is geometrically estimated by using Eq. (3b) (Table 5). The geometrical properties of the landslide in the study area such as length and width of scar, slope angle, thickness

Table 4 Resistance factor assigned to the indicators in the study

Indicators (i)	Category notations	Categories	Raw score	Resistance factor (ξ_i)	Basis for assigning raw score (Literature/IS-Codes)
Year of construction ($\mathbf{R_1}$)	R11	≤ 1990	1	0.33	Puissant et al. (2013), Uzielli et al. (2015)
	R12	1991–2000	2	0.67	
	R13	2001-present date	3	1	
	R14	Unknown	1	0.33	
Roof design ($\mathbf{R_2}$)	R21	Flat roof in wood/bamboo	1	0.33	IS 14804-(2000)
	R22	Pitched roof (slate/local material)	1	0.33	
	R23	Flat/pitched or gable roof in GI or asbestos sheet	2	0.67	
	R24	Flat roof in reinforced concrete (RC)	3	1	
	R25	Pitched or gable roof (IS 14804-(2000), Clause: 6.4.1 & 6.4.2)	3	1	
Structural typology ($\mathbf{R_3}$)	R31	Stone masonry	1	0.25	Uzielli et al. (2015), Silva and Pereira (2014), Li et al. (2010)
	R32	Unreinforced brick masonry	2	0.5	
	R33	Mixed (flat slab with no beam or beam on only one of the floors of building)	3	0.75	
	R34	RC frame with infill walls	4	1	
Surrounding wall of building ($\mathbf{R_4}$)	R41	No/low surrounding wall	1	0.33	Papathoma-Köhle et al. (2007)
	R42	Medium wall	2	0.67	
	R43	Strong high wall	3	1	
Number of floors ($\mathbf{R_5}$)	R51	Single storey (1 storey)	1	0.25	Silva and Pereira (2014), Li et al. (2010), Papathoma-Köhle et al. (2007)
	R52	Low-rise building (2 storeys)	2	0.5	
	R53	Medium-rise building (3, 4 storeys)	3	0.75	
	R54	High-rise building (5, 6 + storeys)	4	1	
Vertical configuration ($\mathbf{R_6}$)	R61	Irregular	1	0.5	IS 14804: (2000) (Clause 6.3.2)
	R62	Regular	2	1	

Table 4 (continued)

Indicators (i)	Category notations	Categories	Raw score	Resistance factor (ξ_i)	Basis for assigning raw score (Literature/IS-Codes)
Horizontal configuration ($\mathbf{R_7}$)	R71	Irregular	1	0.5	IS 14804-(2000); IS 4326: (2013) (Clause 4.4.3)
	R72	Regular	2	1	
Distance between windows & doors from each other: ≥ 1 m ($\mathbf{R_8}$)	R81	Yes	2	1	IS 14804-(2000)
	R82	No	1	0.5	
Distance between two houses from each other: ≥ 1 m ($\mathbf{R_9}$)	R91	Yes	2	1	IS 14804-(2000)
	R92	No	1	0.5	
Building position from steep slope ($\mathbf{R_{10}}$)	R101	Greater than 10 m	3	1	IS 14804-(2000)
	R102	Less than 10 m	2	0.67	
	R103	On the steep slope	1	0.33	
Surrounding vegetation ($\mathbf{R_{11}}$)	R111	Barren	1	0.33	Kappes et al. (2012)
	R112	Cultivated/sparsely vegetated	2	0.67	
	R113	Forest	3	1	
Quality of construction ($\mathbf{R_{12}}$)	R121	Very low	1	0.25	Godfrey et al. (2015)
	R122	Low	2	0.5	
	R123	Medium	3	0.75	
	R124	High	4	1	
State of maintenance ($\mathbf{R_{13}}$)	R131	Poor	1	0.25	Silva and Pereira (2014), Li et al. (2010), Puissant A et al. 2013
	R132	Medium	2	0.5	
	R133	Good	3	0.75	
	R134	Excellent	4	1	

Natural Hazards

Table 4 (continued)

Indicators (*i*)	Category notations	Categories	Raw score	Resistance factor (ξ_i)	Basis for assigning raw score (Literature/IS-Codes)
Type of construction ($\mathbf{R_{14}}$)	R141	Non-engineered	1	0.33	IS 14804-(2000)
	R142	Marginally engineered	2	0.67	
	R143	Engineered	3	1	
Retrofitted building ($\mathbf{R_{15}}$)	R151	Yes	2	1	–
	R152	No	1	0.5	

Resistance Factor: 0 (less resistant) to 1 (maximum resistant)

Raw score is attributed according to the local situation of the study area

Table 5 Empirical relationship for estimating volume of debris flow

Equation no.	Formula	Description of notations	References
3 (a)	$V_L = 0.074 \times A_L^{1.450}$	V_L: volume; and A_L: landslide area	Guzzetti et al. (2009)
3 (b)	$V = A \times h$	V: volume; A: measured area of failed/scoured region; and h: average thickness	Kang and Kim (2016)
3 (c)	$V_{tot} = \frac{1}{6}\pi D_r W_r L_r$	V_{tot}: total volume; D_r: depth of the rupture surface; W_r: width of the dormant landslide; and L_r: length of the dormant landslide	Beyer (1987), Cruden (1990)
3(d)	$V = 0.05 \times S^{1.3}$	V: volume; and S: area	Nikolaeva et al. (2014)

were calculated using Google Earth Image and intensive field surveys (including visual estimation for inaccessible region).

3.5.1.2 Landslide expected velocity (V_{EL}) Determination of landslide velocity is one of the important parameters for the assessing the intensity of a landslide. Cruden and Varnes (1996) presented a range of landslide velocity. Many numerical (continuous and discontinuous methods) and empirical approaches also exist for the estimation of velocity of landslide. There are number of empirical relationships from the scientific literature as shown in Table 6 to estimate debris flow velocity (Chow 1959; Rickenmann 1999; Kang and Kim 2016). In this paper, Eq. (4g) is used to estimate debris flow velocity.

$$V_{\mathrm{EL}} = 2.1 Q_P^{0.33} S^{0.33} \tag{4g}$$

where peak discharge (bouldery material), $Q_P = 0.1 V_{\mathrm{OL}}^{0.83}$ (Rickenmann 1999); and local slope, S = change in elevation (Δh)/horizontal distance (L).

3.5.2 Resistance (*R*) of buildings

Resistance of a building reflects the structural integrity or capacity of building that enables it to resist deformation and destruction during any hazard. The resistance factor of a building not only depends on its structural features but also on the geographical position, workmanship (i.e. quality of construction), architectural features etc.

To determine the resistance ability of buildings in the study area, a cluster of 40 buildings were surveyed. During the extensive field work, documentation of fifteen indicators was carried out as per the BVA form (Table 2) and later resistance factors were assigned to each indicator according to Table 4. Then, the resistance of individual building was calculated using Eq. 5 (Li et al. 2010):

$$R = \left(\prod_{i=1}^{n_s} \xi_i \right)^{1/n_s} \tag{5}$$

where R is the resistance ability of the building, and ξ_i is the ith resistance factor of a particular indicator in case of total number of indicators, $n_s \geq 1$. Each term in this equation represents different vulnerability indicators (expressed non-dimensionally).

Table 6 Empirical relationship for estimating velocity of debris flow

Equation no.	Formula	Description of notations	References
4(a)	$v = \left(\frac{R_c g}{B} \Delta h\right)^{0.5}$	v: velocity; R_c: centre line radius of the bend; g: acceleration due to gravity; B: observed width of flow; and $\Delta h = y_2 - y_1$ (where y_2 and y_1 are the depths at the inner and outer boundaries of the channel)	Henderson (1992) and Chow (1988)
4(b)	$V = \gamma S h^2/(kv)$	γ : unit weight of debris flow; S: slope, h: flow depth; k : cross-sectional shape of coefficient; v: apparent dynamic viscosity of debris	Hungr et al. (1984)
4(c)	$v(x) = \frac{1}{n_c} h^{\frac{2}{3}}(x) I^{\frac{1}{2}}$	n_c: roughness of bed surface; I: slope gradient of the channel; and $h(x)$: flow depth, and equates to $\varphi_2(x) - \varphi_1(x)$	Rickenmann (1999), Chen et al. (2011), Han et al. (2015)
4(d)	$V = (2gh)^{0.5}$	V: estimated mean flow velocity; g: acceleration due to gravity; and h: height of the run-up	Chow (1959)
4(e)	$V = \left[(gdr)/b\right]^{0.5}$	V: estimated mean flow velocity g: acceleration due to gravity; d: super-elevation of the flow; r: radius of the central streamline of the flow, derived from a 1:5,000 scale topographic map; and b: channel width	Chow (1959)
4(f)	$v = a \times h^b S^c$	v: velocity; a: factor that incorporates debris and channel properties; h: flow depth; S: sine of the channel angle, and b and c: exponents that depend on the assumed flow characteristics	Prochaska et al. (2008)
4(g)	$V = 2.1 Q_p^{0.33} S^{0.33}$	V: flow velocity Q_p : peak discharge; S: local slope (S = change in elevation/horizontal distance)	Kang and Kim (2016)

Table 7 Percentage impact of proximity of buildings to landslides at different intensity

S. no.	Proximity of buildings to landslide	High and very high intensity: % decrease in PV	Low and moderate intensity: % decrease in PV
1	≤ 50 m (Very near)	0	0
2	51–100 m (near)	10	5
3	101–150 m (Far)	20	10
4	≥ 151 m (Very far)	30	15

Further, the resistance ability of the buildings in the landslide-prone area has been classified into three classes as: *Low resistanc*e (when $R \leq 0.5$), *moderate resistance* (when $0.5 < R \leq 0.9$) and *high resistance* (when $0.9 < R \leq 1$). This classification scheme will highlight the category of resistance (low, medium, high) of the buildings in a particular built-up region. This will also help in planning the mitigation measures (for existing buildings) in the form of repair and retrofitting to improve their resistances. However, for those buildings which are under construction or will be constructed, new guidelines must be developed to overcome the problems faced earlier.

3.5.3 Proximity of buildings to landslides

A new concept of proximity of buildings to landslide and its impact has been introduced through this study. Kanungo et al. (2008) along with other reasons stated that vulnerability of landslide depends on the resources at risk, their nature and the proximity to the landslide. Since no concrete literature is available in this context, a range of distance (very near i.e. ≤ 50 m to very far i.e. ≥ 151 m) of buildings from landslides with varying intensity has been assumed according to field condition of the study area (Table 7). Here, it is assumed that when the landslide intensity is high and very high, the impact will be high on the buildings in near proximity and the impact will reduce with the increase in the distance of the buildings from the landslide. However, in case of low and moderate landslide intensity, the impact will be comparatively less. Therefore, according to the field condition and expert knowledge, a decrease of the order of 0 to 30% in PV (as obtained by Eq. 6) depending on the very near to very far proximity of buildings to the landslides with high and very high intensities and a decrease of the order of 0 to 15% in PV in case of landslides with low and moderate intensities are assumed.

3.6 Quantification of vulnerability at site-specific scale

In the literature, a number of frameworks either quantitative or semi-quantitative have been proposed to assess physical vulnerability of buildings. For instance, in the research conducted by Papathoma-Köhle et al. (2007), five indicators were used to assess the vulnerability of each building and the weights were assigned to each indicator according to the priorities and the final purpose of assessment. Major limitation of the study was that the magnitude and frequency of landslide was not taken into consideration. Only four indicators were used in the method proposed by Li et al. (2010). However, information on some indicators like debris-depth, deformation factor of structure, type of foundation soil and the depth of foundation are not always available for estimation of vulnerability of buildings. Most of the research works

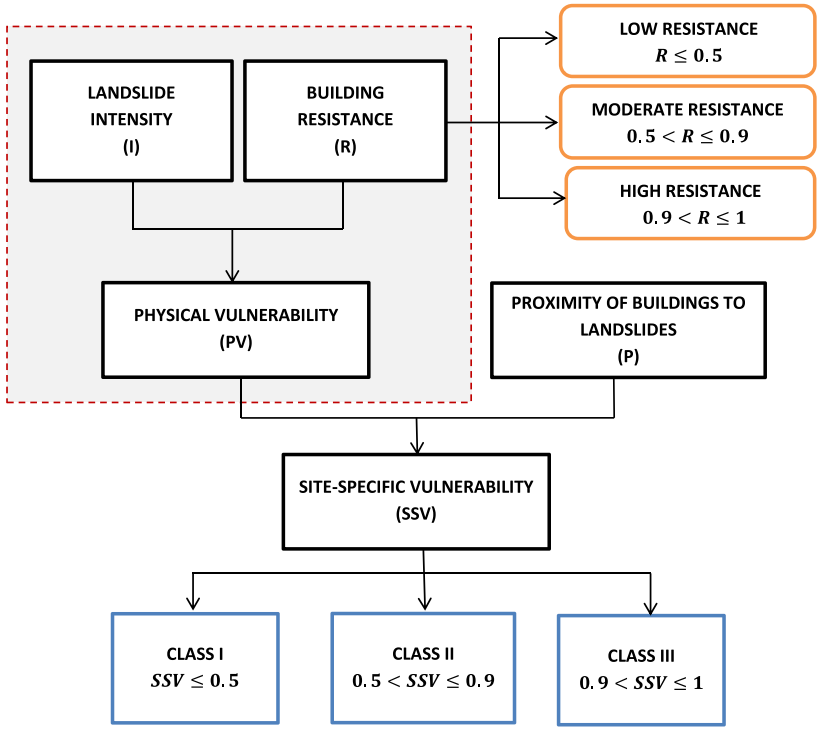


Fig. 1 Flowchart of the methodology applied in the study

carried out in the past are scenario-based conceptual framework (Papathoma-Köhle et al. 2007; Uzielli et al. 2008; Li et al. 2010; Kappes et al. 2012) and only few have been applied at regional scale in practice. Also, the previous proposed works are specific only to European countries and incorporate a maximum of four to five indicators for assessing the physical vulnerability of buildings. Some studies lack information on the intensity of landslide, while some studies have ranked the indicators describing that one indicator contributes more in vulnerability than others. It is the aim of this study to fill the gap and make the method more user-friendly.

The flowchart of methodology applied in this study is shown in Fig. 1. The methodology represents quantification of physical vulnerability (PV) based on landslide intensity (I) and resistance ability (R) of the building. PV in this study is quantified using the empirical relationship (Eq. 6) proposed by Li et al. (2010).

$$PV = f(I,R) = \begin{cases} 2\frac{I_s^2}{R^2} & \frac{I_s}{R} \leq 0.5 \\ 1 - \frac{2(R-I_s)^2}{R^2} & 0.5 < \frac{I_s}{R} \leq 1.0 \\ 1.0 & \frac{I_s}{R} > 1 \end{cases} \quad (6)$$

here I_s is the landslide intensity observed in the study area (since the landslide intensity might vary from one region to another).

Further, the buildings which are surrounding the landslide will consequently suffer more damage in comparison to the building which is away from the landslide zone. Thus, vulnerability of the building also tends to change as per its proximity to the landslide. If the landslide zone is near to built-up area, the buildings will be more vulnerable and vice versa. Therefore, proximity (P) of landslide to building is an essential parameter in assessing site-specific vulnerability (SSV) in the affected region.

Therefore, depending upon the type of intensity and its distance from landslide zone, the site-specific vulnerability of building will be estimated using Eq. 7.

$$SSV = PV - P \quad (7)$$

where PV is physical vulnerability of building and P is the per cent decrease in PV due to proximity.

In order to understand the degree of vulnerability of the buildings (describing the damage pattern), SSV is classified into three classes (Class I, Class II and Class III) on a ratio scale as proposed by Singh et al. (2018). Class I ($SSV \leq 0.5$), Class II ($0.5 < SSV \leq 0.9$) and Class III ($0.9 < SSV \leq 1.0$) exhibits low damage, moderate to severe damage and very severe to complete damage, respectively. The details of all the three classes as given in Singh et al. (2018) are as follows:

(i) Class I: This means, the buildings in this class will have very less chance of damage, where the functionality of the building is not compromised and the minor damage can be repaired at very nominal cost.
(ii) Class II: In this class, the functionality of the building is compromised and immediate evacuation of the occupants is necessary in order to prevent further accidents due to falling masonry. The repair and retrofitting of such buildings will involve higher cost and manpower.
(iii) Class III: Such type of damage cannot be fixed and requires clearance of the site as the stability of the building is compromised.

4 A case study

4.1 Study area

Uttarakhand lies in a seismically active region (zone IV and V) and also shares a long history of landslide events such as Malpa landslide (1998), Varunavat Parvat landslide (2003), the Rishikesh-Badrinath National Highway at Chhinka in Chamoli district (29 July 2010) and the landslides initiated due to 2013 flash-flood events in Kedarnath. Therefore, an area where a community is prone to landslide has been selected for this study.

Gopeshwar Township of Chamoli district is taken as the study area for assessing the vulnerability of buildings exposed to landslides (Fig. 2). The area lies in the north-eastern part of Uttarakhand state (latitudes 30°25″N–79°20″E and longitudes 30.42°N–79.33°E). Gopeshwar is the district headquarters of Chamoli district (Uttarakhand), perched at an elevation of 1550 m. This town shares its northern boundary with Tibet (China) and eastern boundary with Nepal. The whole district is drained by Alaknanda River which is one of the main sources of the river Ganges. The study area which is located in the upper Alaknanda valley is highly prone particularly to debris slide and debris flow type of landslides (Sajwan

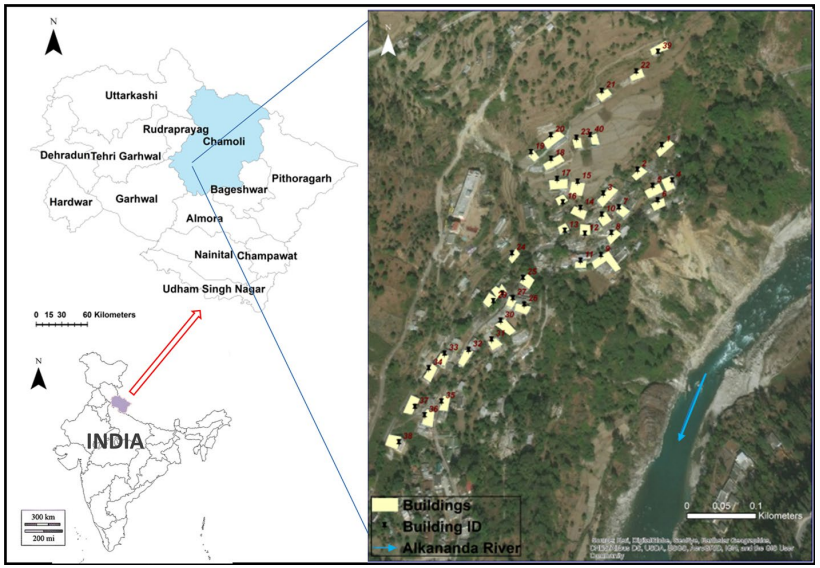


Fig. 2 Location of the study area

and Sushil 2016). Therefore, landslide zones surrounding the Gopeshwar town of Chamoli district (Uttarakhand) pose a serious risk to the buildings, infrastructures and people living in that area.

4.2 Landslide in the study area

Gopeshwar Township has been divided into 11 wards as per Municipality (Nagar Palika Parishad, Chamoli). The particular landslide undertaken in this study lies in ward no. 10 (Kothiyal Sain) of Gopeshwar Township and covers an area of 1.55 square kilometres having around 459 houses.

The landslide is located at 30°23′48.29″N latitude and 79°19′19.94″E longitude. The landslide appears like a debris slide from the face due to gully erosion by the drainage channels flowing from the uphill slope. However, the source area for possible debris flow along these drainage channels with landslide scars is clearly visible on the uphill slope beyond the habitated area (building clusters). A number of drainage channels are connected from the source area to the Alaknanda River and they pass through the habitated area and the debris slide on the riverside. Some of these drainage channels have already witnessed debris flow to some extent (Fig. 3). During the field mapping, it was observed that beyond the right flank of slide along drainage channels check dams have also been constructed as a mitigation measure, to reduce further flow of debris material. However, since the landslide is active, the first portion of the check dam has moved from its original position. The face of the landslide zone is 137 m long and 299 m wide and encompasses an area of about 28,195 m^2. The scare at the uphill side (i.e. source zone of landslide) is 102 m long and 298 m wide with an average thickness of 0.6 m and covers an area of about

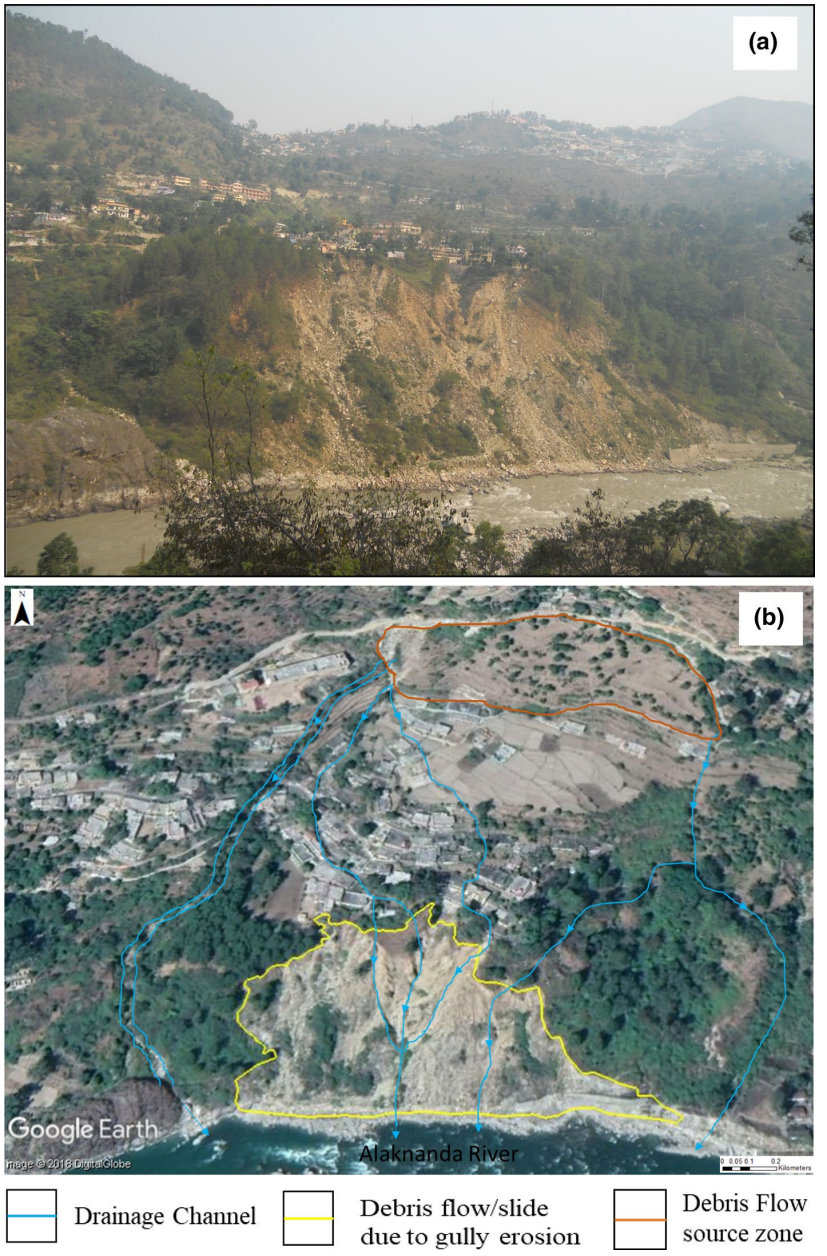


Fig. 3 Debris flow in ward no. 10 (Kothiyal Sain) of Gopeshwar Township, Chamoli District (Uttarakhand) shown in **a** field photograph and **b** Google Earth image

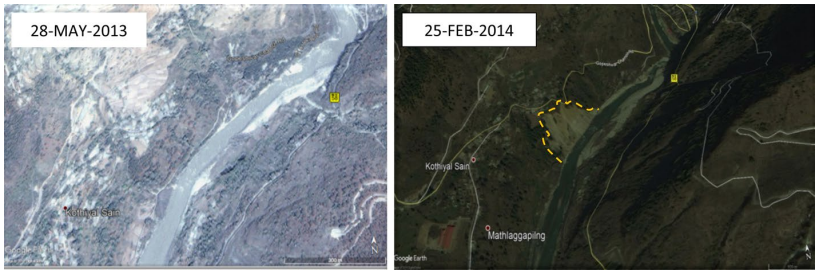


Fig. 4 Comparison of Google Earth image in the study area for the year 2013 and 2014

29,626 m^2. Geomorphologically, the slope is steep with slope angle varying between 60° and 75° and the terrain is rugged. The vegetation is uniform in the area except on the main landslide body. The tilted trees indicated that the slide is progressive and still active in nature. According to local people, the landslide occurred during the 2013 flash floods (in rivers Bhagirathi, Alaknanda, Mandakini, Dhauliganga and Kali) between June 15 and 17, 2013 incident and the affected area used to have dense vegetation cover. This information was validated by studying the historical imagery from Google Earth for the year 2013 and 2014 (Fig. 4).

Just above the landslide zone lies densely populated built-up area. On comparing the images (Google Earth) from the years 2014 and 2017 (Fig. 5), it is evident that a number of construction activities took place in the surrounding uphill slopes of the landslide zone. Field survey indicated that the region (ward no. 10) is under subsidence due to this landslide along with possibility of debris flow along drainage channels. Figure 6 shows a school building which is located just above the landslide (sharing its boundary). The school (Building ID 9) is worst affected in the whole ward as the cracks not only on walls but also on the floor can be clearly seen; in fact a portion of the school is lost when the landslide occurred.

4.3 Observations on the buildings in the study area

The buildings constructed in the hilly terrain of India are significantly different from other Indian subcontinent. During extensive field surveys of buildings in ward no. 10 of Gopeshwar Township, the following features have been observed:

(i) Most of the buildings were non-engineered or marginally engineered;
(ii) Due to the unavailability of flat lands in hilly terrain, most of the buildings have foundations at different levels. Some buildings have number of storeys above road level, while some are constructed below road level with entrance levels towards the slope or against the slope.
(iii) Traditional houses made of random rubble stone masonry were also found in the study area.
(iv) A number of drainage channels starting from the landslide scars (source area) on the uphill slope are passing through the clusters of buildings on the upper side of the debris slide on the face due to gully erosion by these drainage channels. Hence, there is a possibility of debris flow along these channels as one of these channels has already witnessed debris flow. Thus, the buildings in this area are under different degrees of vulnerability due to such debris flow. Such instability in the area has also been witnessed in one of the newly constructed buildings with development of cracks (refer Fig. 7).

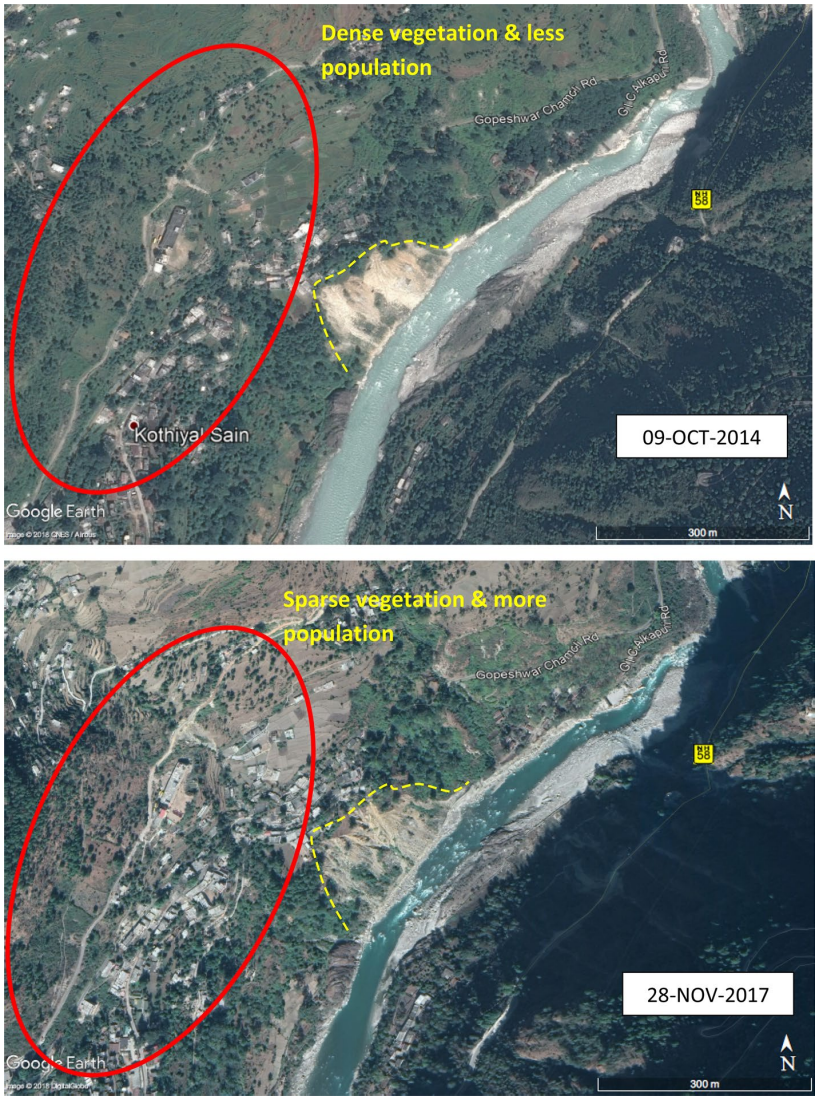


Fig. 5 Google Earth images showing increase in construction activities around debris flow in the study area from the year 2014 & 2017



Fig. 6 Damage caused to school building (building ID: 9) near landslide zone (red arrow showing cracks at various places)

5 Results and discussions

In the present study, the vulnerability of buildings is determined on the basis of the landslide intensity (I), the resistance ability of the buildings (R) and the proximity of buildings (P) from debris flow zone of influence (i.e. the landslide scar in the source area, the drainage channels and the debris slide on riverside as an effect of gully erosion). The methodology has been implemented at a site-specific scale for a single debris flow site at Gopeshwar Township located in Chamoli district of Uttarakhand (India). The results obtained after implementing the methodology are discussed in the following section.

5.1 Landslide intensity (I)

The volume of debris flow in the study area estimated using Eq. 3b (Table 5) is of the order of 0.02×10^6 m^3. Therefore, the estimated landslide volume falls in the classification range of 'high' as proposed by Sarkar et al. 2015. Further, velocity of the debris flow is estimated using the empirical relationship (Eq. 4g) developed by Rickenmann (1999) as given in Table 6. Since the debris is comprised of bouldery material, peak discharge is estimated using suitable relationship. The estimated velocity obtained is 2.3 m/s. According to the velocity scale proposed by Cruden and Varnes (1996), the estimated velocity obtained is in between 5 mm/s to 5 m/s, corresponding to very rapid type of flow. Based on the landslide intensity matrix as proposed by Singh et al. (2018), the landslide intensity in this case is

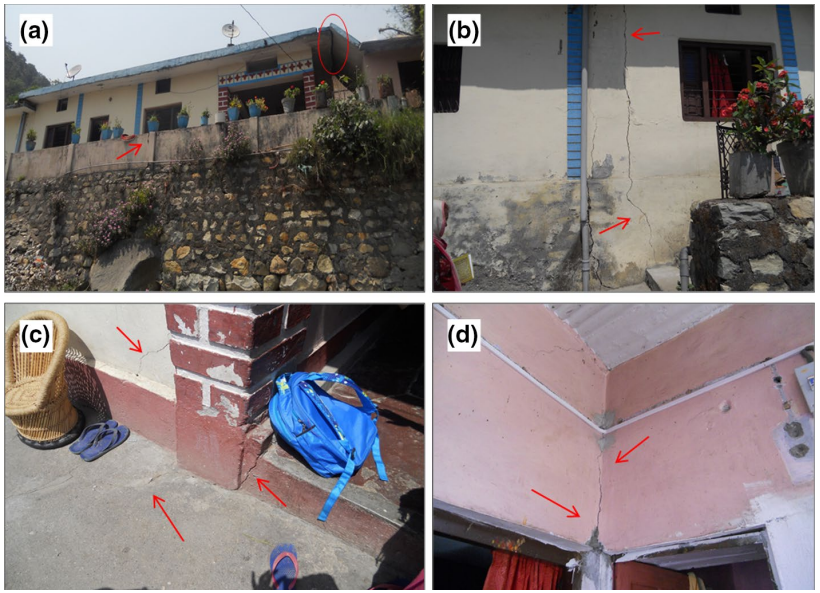


Fig. 7 Residential building (building ID: 12) in the study area showing cracks (red arrow) at various places: **a** surrounding wall of the building and on roof joint, **b** exterior wall, **c** wall and floor and **d** at corner of wall

obtained to be very high; hence, the landslide intensity weightage value is taken as 1.0 (i.e. $I = 1$).

5.2 Resistance of building (R) obtained in the study area

Every physical and geographical aspect of a building play a significant role in strengthening the whole building in order to avoid damage during disasters like landslide (specifically debris flow as in the present case). Therefore, the only way to protect the buildings and infrastructure in landslide-prone region is by increasing the resistance ability of the buildings and minimising the intensity of landslide. In order to achieve this objective, first the information of individual building in the study area is recorded in BVA form during an extensive field survey. The percentage distribution of the resistance ability offered by each category of all the fifteen vulnerability indicators considered in this study is presented in Fig. 8. It is observed that maximum number of buildings in the study area are low-rise building constructed in the year prior to 1990, having mixed structural typology and flat roof in reinforced concrete (RC) with no/low surrounding wall. The buildings constructed in the study area as per the IS (Indian standard) codes are rare, since 68% buildings have irregular vertical configuration, 65% buildings have irregular horizontal configuration, 2% buildings have distance > 1 m between doors and windows and only 2% of buildings have distance > 1 m between two houses from each other. Maximum number of buildings have medium quality of construction and medium state of maintenance. It is also observed that 30% of the buildings are non-engineered, while 61% are marginally engineered. The

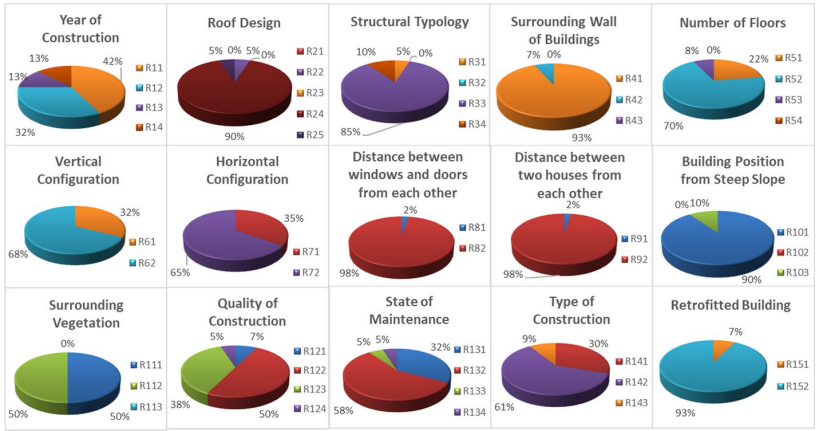


Fig. 8 Percentage distribution of different categories of vulnerability indicators of the buildings under study (refer Table 4 for explanation of R11 to R152)

recorded vulnerability indicators and the final resistance of the building is summarised in Table 8. It can be observed that the minimum and maximum resistance offered by the building in the study area is 0.43 (Building ID 9) belonging to low resistance ($R \leq 0.5$) and 0.73 (Building ID 3 and 38) belonging to moderate resistance category ($0.5 < R \leq 0.9$), respectively. Therefore, from the obtained results (Table 8), it can be summarised that 7.5% of buildings provide low resistance and 92.5% of buildings provide moderate resistance. Since the buildings in the study area mostly belong to the people with poor economic background, these have low and moderate resistance ability. Also, most of the buildings in hilly terrains are non-engineered and the masons lack sufficient knowledge regarding landslide resistant construction techniques; therefore, none of the buildings provided high resistance ability.

5.3 Physical vulnerability (PV) and Site-specific vulnerability (SSV) of buildings

As discussed in Sect. 3.6, physical vulnerability (PV) of every individual building is calculated based on landslide intensity (as discussed in Sect. 5.1) and building resistance (as discussed in Sect. 5.2) as per Eq. (6). It is observed that PV for all the buildings in the study area are equal to 1. This means that all the buildings will suffer total damage in the study area due to the possible debris flow scenario. However, the distance of the building from the debris flow source area and influence zone also plays an important role in assessing the vulnerability. Therefore, proximity parameter as discussed in Sect. 3.5.3 has been considered in calculating the site-specific physical vulnerability (SSV) for each building in the study area (as given in Table 9). As shown in the flowchart (Fig. 1), the SSV values thus obtained are then classified into different vulnerability classes (as provided in Sect. 3.6). Figure 9 shows the distribution of buildings in the study area into different SSV classes. It is observed that 68% of the buildings fall under Class III ($0.9 < \text{SSV} \leq 1.0$) and the rest 32% of buildings fall under class II. Figure 10 shows the spatial pattern of site-specific vulnerability of buildings in the study area falling under class II and class III.

Table 8 Resistance (R) of the buildings in the study area

Building ID	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R
1	1	1	0.75	0.33	0.5	0.5	0.5	0.5	0.5	1	0.33	0.5	0.5	0.33	0.5	0.54
2	0.33	1	0.75	0.33	0.5	1	0.5	0.5	0.5	1	0.33	0.5	0.5	0.33	0.5	0.53
3	0.67	1	0.75	0.67	0.5	1	1	1	1	1	0.67	0.5	0.5	0.67	0.5	0.73
4	0.33	0.33	0.25	0.33	0.25	1	1	0.5	0.5	1	0.67	0.75	0.5	0.33	1	0.51
5	0.67	1	0.75	0.67	0.5	1	1	0.5	0.5	1	0.33	1	1	0.67	0.5	0.70
6	0.67	1	0.75	0.33	0.75	0.5	1	0.5	0.5	0.33	0.33	0.5	0.25	0.67	0.5	0.53
7	0.67	1	0.75	0.33	0.25	0.5	1	0.5	0.5	0.33	0.33	0.5	0.25	0.67	0.5	0.49
8	1	1	0.75	0.33	0.25	0.5	1	0.5	0.5	0.33	0.33	0.75	0.5	0.67	0.5	0.54
9	0.67	1	0.75	0.33	0.25	0.5	0.5	0.5	0.5	0.33	0.33	0.25	0.25	0.33	0.5	0.43
10	0.33	1	0.75	0.33	0.25	0.5	1	0.5	0.5	1	0.33	0.25	0.25	0.67	1	0.50
11	0.33	1	1	0.33	0.25	0.5	0.5	0.5	0.5	1	0.67	0.5	0.5	0.67	0.5	0.54
12	0.33	1	0.75	0.33	0.5	0.5	1	0.5	0.5	1	0.67	0.5	0.25	0.33	0.5	0.53
13	0.33	1	0.75	0.33	0.5	0.5	1	0.5	0.5	1	0.67	0.25	0.5	0.33	0.5	0.53
14	0.33	1	0.75	0.33	0.25	0.5	1	0.5	0.5	1	0.67	0.5	0.5	0.67	0.5	0.55
15	0.33	1	0.75	0.33	0.5	1	1	0.5	0.5	1	0.67	0.5	0.25	0.33	0.5	0.55
16	0.33	1	1	0.33	0.75	1	0.5	0.5	0.5	1	0.33	1	1	0.67	0.5	0.64
17	0.67	1	0.75	0.33	0.5	1	1	0.5	0.5	1	0.33	0.75	0.5	0.33	0.5	0.60
18	0.67	1	0.75	0.33	0.5	1	1	0.5	0.5	1	0.33	0.75	0.5	0.67	0.5	0.62
19	0.67	1	1	0.33	0.5	1	1	0.5	0.5	1	0.33	0.75	0.5	0.67	0.5	0.64
20	0.33	1	0.75	0.33	0.5	1	0.5	0.5	0.5	1	0.67	0.5	0.5	0.67	0.5	0.58
21	1	1	1	0.33	0.5	1	1	0.5	0.5	1	0.67	0.75	0.5	0.33	0.5	0.65
22	1	1	0.75	0.33	0.5	1	1	0.5	0.5	1	0.67	0.75	0.75	0.67	0.5	0.69
23	1	1	0.75	0.33	0.75	0.5	0.5	0.5	0.5	1	0.67	0.75	0.75	0.67	0.5	0.65
24	0.33	1	0.75	0.33	0.5	1	0.5	0.5	0.5	1	0.33	0.5	0.5	0.33	0.5	0.53
25	0.33	1	0.75	0.33	0.5	1	0.5	0.5	0.5	1	0.67	0.75	0.25	0.33	0.5	0.54
26	0.33	1	0.75	0.33	0.5	1	0.5	0.5	0.5	1	0.67	0.5	0.5	0.67	0.5	0.58
27	0.67	1	0.75	0.33	0.5	1	1	0.5	0.5	1	0.33	0.75	0.5	0.33	0.5	0.60

Table 8 (continued)

Building ID	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R
28	0.33	1	0.75	0.33	0.5	0.5	1	0.5	0.5	1	0.67	0.5	0.25	0.33	0.5	0.53
29	0.33	1	0.75	0.33	0.5	1	1	0.5	0.5	1	0.67	0.5	0.25	0.33	0.5	0.55
30	0.67	1	0.75	0.33	0.5	1	1	0.5	0.5	1	0.33	0.75	0.5	0.67	0.5	0.62
31	0.33	0.33	0.25	0.33	0.25	1	1	0.5	0.5	1	0.67	0.75	0.5	0.33	1	0.51
32	0.67	1	0.75	0.33	0.5	1	1	0.5	0.5	1	0.33	0.75	0.5	0.33	0.5	0.60
33	0.33	1	0.75	0.33	0.5	1	0.5	0.5	0.5	1	0.67	0.5	0.5	0.67	0.5	0.58
34	0.67	1	0.75	0.33	0.5	1	1	0.5	0.5	1	0.33	0.75	0.5	0.67	0.5	0.62
35	0.33	1	0.75	0.33	0.5	1	0.5	0.5	0.5	1	0.33	0.5	0.5	0.33	0.5	0.53
36	0.33	1	0.75	0.33	0.5	1	0.5	0.5	0.5	1	0.67	0.75	0.25	0.33	0.5	0.54
37	0.33	1	0.75	0.33	0.5	1	1	0.5	0.5	1	0.67	0.5	0.25	0.33	0.5	0.53
38	0.33	1	0.75	0.33	0.5	1	0.5	0.5	0.5	1	0.33	0.5	0.5	0.33	0.5	0.73
39	0.67	1	0.75	0.33	0.25	0.5	1	0.5	0.5	1	0.33	0.5	0.25	0.67	0.5	0.51
40	0.33	1	0.75	0.33	0.5	1	1	0.5	0.5	1	0.67	0.5	0.25	0.33	0.5	0.70

Table 9 Site-Specific Vulnerability (SSV) of the buildings in the study area

Building ID	Landslide intensity (I_s)	Resistance, (R)	Physical vulnerability (PV) (as per Eq. 2)	Proximity of building to landslide (P)	Site-specific vulnerability (SSV) (as per Eq. 3)	SSV class
1	Very high (1.0)	0.54	1	≤ 50 m	1	III
2		0.53	1	≤ 50 m	1	III
3		0.73	1	≤ 50 m	1	III
4		0.51	1	≤ 50 m	1	III
5		0.70	1	≤ 50 m	1	III
6		0.53	1	≤ 50 m	1	III
7		0.49	1	≤ 50 m	1	III
8		0.54	1	≤ 50 m	1	III
9		0.43	1	≤ 50 m	1	III
10		0.50	1	≤ 50 m	1	III
11		0.54	1	≤ 50 m	1	III
12		0.53	1	≤ 50 m	1	III
13		0.53	1	≤ 50 m	1	III
14		0.55	1	≤ 50 m	1	III
15		0.55	1	≤ 50 m	1	III
16		0.64	1	≤ 50 m	1	III
17		0.60	1	≤ 50 m	1	III
18		0.62	1	≤ 50 m	1	III
19		0.64	1	≤ 50 m	1	III
20		0.58	1	≤ 50 m	1	III
21		0.65	1	≤ 50 m	1	III
22		0.69	1	≤ 50 m	1	III
23		0.65	1	≤ 50 m	1	III
24		0.53	1	≤ 50 m	1	III
25		0.54	1	≤ 50 m	1	III
26		0.58	1	101–150 m	0.8	II
27		0.60	1	≥ 151 m	0.7	II
28		0.53	1	≥ 151 m	0.7	II
29		0.55	1	≥ 151 m	0.7	II
30		0.62	1	≥ 151 m	0.7	II
31		0.51	1	≥ 151 m	0.7	II
32		0.60	1	≥ 151 m	0.7	II
33		0.58	1	≥ 151 m	0.7	II
34		0.62	1	≥ 151 m	0.7	II
35		0.53	1	≥ 151 m	0.7	II
36		0.54	1	≥ 151 m	0.7	II
37		0.53	1	≥ 151 m	0.7	II
38		0.73	1	≥ 151 m	0.7	II
39		0.51	1	≤ 50 m	1	III
40		0.70	1	≤ 50 m	1	III

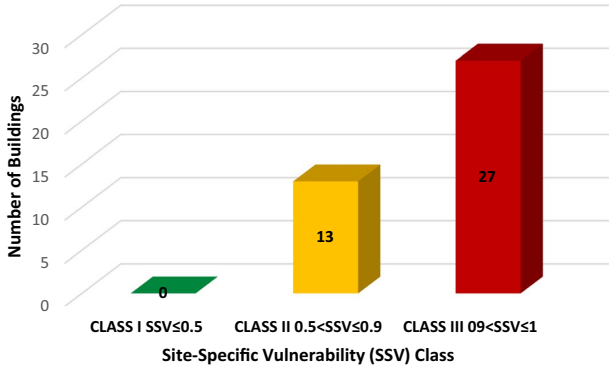


Fig. 9 Histogram showing distribution of buildings in different classes of site-specific vulnerability (SSV) in the study area

6 Concluding remarks

Despite widespread interest on landslide research in India, very few published studies are found on physical vulnerability of buildings exposed to landslides. Moreover, unlike the earthquake resistant codes on design of structures in India, no proper guidelines exist on the design of buildings in landslide-prone region. In this paper, a simple and user-friendly approach that can be easily implemented on ground level is presented to assess the vulnerability of buildings exposed to landslides in India. Our work suggests that the vulnerability of a building exposed to landslides can be assessed based on the landslide intensity, resistance of the buildings and the proximity of the buildings from the landslide-affected region. Landslide intensity in the study area is estimated using well-established empirical relationships. The key elements in this research are the identification of fifteen vulnerability indicators (that include structural, architectural, topographic, geographical and workmanship parameters) for assessing the resistance ability of buildings and the proximity of buildings from landslide influence zone.

The vulnerability indicators selected in this study exhibit certain degree of building resistance that prevents damage or collapse of building during landslide. Different categories within each indicator have been assigned weightages on the scale of 0 (i.e. least resistance) and 1 (i.e. most resistance), depending upon their probable resistance ability in case of landslide exposure. In this study, the study area is exposed to very high landslide intensity (i.e. $I=1$) due to a debris flow having 7.5% of low resistance buildings and 92.5% of moderate resistance buildings. However, through this method individual resistance ability (low, moderate and high resistance) of the buildings can also be identified. This categorisation will help in prioritising the buildings that requires remedial measures such as retrofitting to minimise the impact of destruction during the landslide event.

Further, proximity of the buildings from landslide-prone region is relatively a new concept in assessing the site-specific vulnerability of buildings. In this study, the physical vulnerability is estimated for very high intensity landslide (i.e. debris flow), which shows that all the buildings in the study area will fall under class III physical vulnerability (PV), i.e. the occurrence of this debris flow will cause complete damage to the buildings in this ward. However, practically, this scenario looks unrealistic as all the buildings will not face

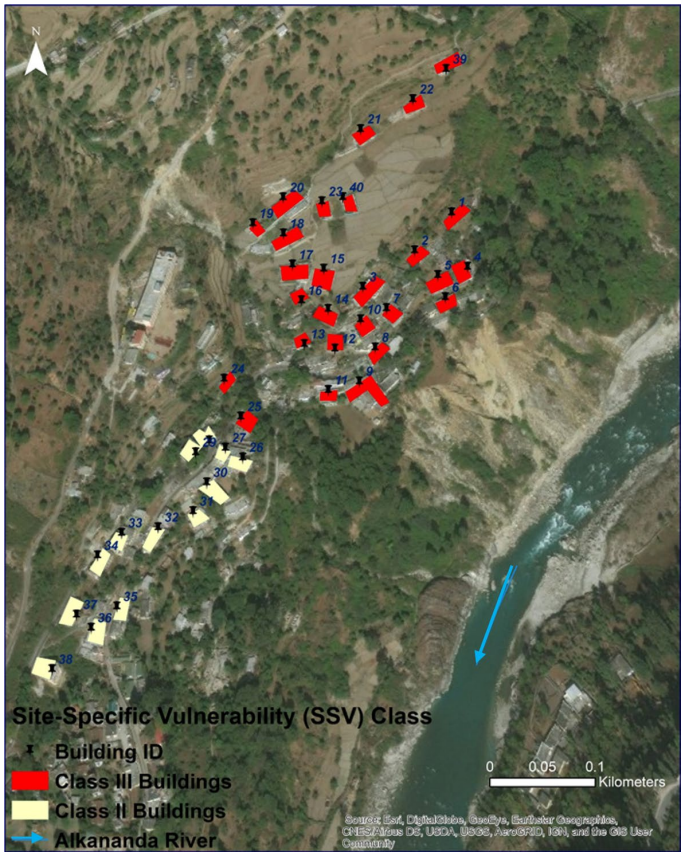


Fig. 10 Spatial pattern of site-specific vulnerability (SSV) classes of buildings in the study area

similar amount of damage due to the debris flow. The physical vulnerability (PV) will vary based on its distance from the influence zone of debris flow. Therefore, a concept of proximity of the buildings from the influence zone of debris flow is introduced to assess the site-specific vulnerability (SSV) of each building in the ward. The ground evidence reflects that closer the building more will be its vulnerability. Hence, this also justifies the fact that the degree of vulnerability of individual building will be over-estimated if the proximity parameter of the building is not considered. In the present work, the inclusion of proximity of the buildings from the landslide-prone zone clearly demonstrates that the buildings constructed around close proximity to the debris flow influence zone fall under high vulnerability class though they have displayed low (e.g. Building ID: 7, 9 and 10) and moderate (e.g. Building ID: 1–6, 8 and 11–40) resistance ability. Those buildings (e.g. Building IDs: 26 to 38) with moderate resistance ability and in far proximity to the debris flow influence zone fall under moderate vulnerability class. Thus, it is observed that 27 out of 40

buildings fall under Class III ($0.9 < SSV \leq 1.0$) and 13 out of 40 buildings fall under class II SSV.

Therefore, it can be concluded that this methodology can be used in a quite realistic manner to assess the vulnerability of buildings exposed to debris flows and similar potential failures. Such assessment can be extensively used by engineers and scientists to facilitate appropriate protective measures to both the buildings and the landslide to increase the resistance and decrease the intensity, respectively, in a particular area. This will reduce the vulnerability of buildings due to landslides and thereby will reduce the risk of loss of lives and property. Such type of studies will further help in preparing the construction guidelines specific to a particular region, which are presently lacking in the country.

Acknowledgements The second author wishes to thank the Director, CSIR-Central Building Research Institute, Roorkee, for his kind permission to publish this work.

References

Aleotti P, Chowdhury R (1999) Landslide Hazard Assessment: summary, review and new perspectives. Bull Eng Geol Environ 58:21–44

Beyer WH (ed) (1987) Handbook of mathematical sciences, 6th edn. CRC Press, Boca Raton, p 860

Birkmann J (2006) Indicators and criteria for measuring vulnerability: theoretical bases and requirements. In: Birkmann J (ed) Measuring vulnerability to natural hazards: towards disaster resilient societies. UNU Press, Tokyo

Birkmann J, Cardona OM, Carreno ML, Barbat AH, Pelling M, Schneiderbauer S, Kienberger S, Keiler M, Alexander D, Zeil P, Welle T (2013) Framing vulnerability, risk and societal responses: the MOVE framework. Nat Hazards 67(2):193–211

Cardinali M, Reichenbach P, Guzzetti F, Ardizzone F, Antonini G, Galli M, Cacciano M, Castellani M, Salvati P (2002) A geomorphological approach to the estimation of landslide hazard and risk in Umbria, Central Italy. Nat Hazards Earth Syst Sci 2:1–16. https://doi.org/10.5194/nhess-2-57-2002

Chen LX, Yin KL, Dai YX (2011) Building vulnerability evaluation in landslide deformation phase. J Mt Sci 8:286–295. https://doi.org/10.1007/s11629-011-2101-z

Chow VT (1959) Open-channel hydraulics. McGraw-Hill, New York, p 680

Chow VT (1988) Open-channel hydraulics. McGraw-Hill, McGraw-Hill classic textbook reissue series

Cooper AH (2008) The classification, recording, data basing and use of information about building damage caused by subsidence and landslides. Q J Eng Geol Hydrogeol 41:409–424

Cruden DM (1990) A suggested method for reporting a landslide. Bull Int As Eng Geol 41:5–12

Cruden DM, Varnes DJ (1996) Landslide types and processes. In: Turner AK, Schuster RL (eds) Landslides: investigation and mitigation (Special Report 247). National Research Council, Transportation and Research Board, Washington, DC, pp 36–75

Dai FC, Lee CF, Ngai YY (2002) Landslide risk assessment and management: an overview. Eng Geol 64(1):65–87

Del Soldato M, Bianchini S, Calcaterra D, De Vita P, Martire DD, Tomas R, Casagli N (2017) A new approach for landslide induced damage assessment. Geomat Nat Hazards Risk 8(2):1524–1537. https://doi.org/10.1080/19475705.2017.1347896

Douglas J (2007) Physical vulnerability modeling in natural hazard risk assessment. Nat Hazards Earth Syst Sci 7:283–288

Du J, Yin K, Lacasse S, Nadim F (2014) Quantitative vulnerability estimation of structures for individual landslide: application to the metropolitan area of San Salvador, El Salvador. Electron J Geotech Eng 19:1251–1264

EPFL: EcolePolytechnique de Lausanne (2002). Relevant Criteria to assess vulnerability and risk. Unpublished Deliverable (D16) of project IMIRILAND: Impact of Large Landslides in the mountain environment

Ettinger S, Mounaud L, Magill C, Yao-Lafourcade A-F, Thouret J-C, Manville V, Negulescu C, Zuccaro G, de Gregorio D, Nardone S, LuqueUchuchoque JA, Arguedas A, Macedo L, ManriqueLlerena N (2016) Building vulnerability to hydro-geomorphic hazards: estimating damage probability from qualitative vulnerability assessment using logistic regression. J Hydrol 541(PART A):563–581

Fuchs S, Heiss K, Hübl J (2007) Towards an empirical vulnerability function for use in debris flow risk assessment. Nat Hazards Earth Syst Sci 7(5):495–506

Geological Survey of India, Ministry of Mines, Government of India, http://www.gsi.gov.in. Retrieved on 7 August 2018

Ghosh S, Carranza EJM, van Westen CJ, Jetten VG, Bhattacharya DN (2011) Selecting and weighting spatial predictors for empirical modeling of landslide susceptibility in the Darjeeling Himalayas (India). Geomorphology 131:35–56. https://doi.org/10.1016/j.geomorph.2011.04.019

Glade T (2003) Vulnerability assessment in landslide risk analysis. Die Erde 134:121–138

Godfrey A, Ciurean RL, van Westen CJ, Kingma N, Glade T (2015) Assessing vulnerability of buildings to hydro-meteorological hazards using an expert based approach - an application in Nehoiu Valley, Romania. Int J Disaster Risk Reduct 13:229–241

Guzzetti F, Ardizzone F, Cardinali M, Rossi M, Valigi D (2009) Landslide volumes and landslide mobilization rates in Umbria, central Italy. Earth Planet Sci Lett 279:222–229

Hamilton LS (2008) A thematic study prepared in the framework of the Global Forest resources Assessment, 2005. FAO Forestry Paper 155. Food and Agriculture Organization of the United Nations, Rome; Lal, M., 2001: tropical cyclones in a warmer world. Curr Sci India 80:1103–1104

Han Z, Chen G, Li Y, Tang C, Xu L, He Y, Huang X, Wang W (2015) Numerical simulation of debris-flow behavior incorporating a dynamic method for estimating the entrainment. Eng Geol 190:52–64

Henderson F (1992) Open channel flow. Gale Group, Farmington Hills

Hungr O, Morgan GC, Kellerhals R (1984) Quantitative-analysis of debris torrent hazards for design of remedial measures. Can Geotech J 21(4):663–677

IS 14804 (2000) Siting, design and selection of materials for residential buildings in hilly areas–guidelines. In: Bureau of Indian Standards, New Delhi, India

IS 1893 - part 1 (2016) Criteria for earthquake resistant design of structures. In: Bureau of Indian Standards, New Delhi, India

IS 4326 (2013) Earthquake resistant design and construction of buildings — code of practice. In: Bureau of Indian Standards, New Delhi, India

ISDR (2009) Global assessment report on disaster risk reduction. United Nations, Geneva

Kang H, Kim Y (2016) The physical vulnerability of different types of building structure to debris flow events. Nat Hazards 80:1475–1493

Kanungo DP, Arora MK, Sarkar S, Gupta RP (2006) A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. Eng Geol 85:347–366

Kanungo DP, Arora MK, Gupta RP, Sarkar S (2008) Landslide risk assessment using concepts of danger pixels and fuzzy set theory in Darjeeling Himalayas. Landslides 5:407–416. https://doi.org/10.1007/s10346-008-0134-3

Kappes M, Papathoma-Köhle M, Keiler M (2012) Assessing physical vulnerability for multi-hazards using an indicator-based methodology. Appl Geogr 32(2):577–590

Li Z, Nadim F, Huang H, Uzielli M, Lacasse S (2010) Quantitative vulnerability estimation for scenario-based landslide hazards. Landslides 7(2):125–134. https://doi.org/10.1007/s10346-009-0190-3

Mathew J, Jha VK, Rawat GS (2009) Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method. Landslides 6:17–26

Menoni S (2006) Integration of harmonized risk maps with spatial planning decision processes. Deliverable 5.1, Armonia

National Building Code of India (2016) Bureau of Indian Standards, vol 1, New Delhi-110002

Nikolaeva E, Walter TR, Shirzaei M, Zschau J (2014) Landslide observation and volume estimation in central Georgia based on L-band InSAR. Nat Hazards Earth Syst Sci 14:675–688

Papathoma-Köhle M (2016) Vulnerability curves versus vulnerability indicators: application of an indicator-based methodology for debris-flow hazards. Nat Hazards Earth Syst Sci 16:1771–1790. https://doi.org/10.5194/nhess-16-1771-2016

Papathoma-Kohle M, Totschnig R, Keiler M, Glade T (2012) A new vulnerability function for debris flow e the importance of physical vulnerability assessment in alpine areas. In: Koboltschng G, Hübl J, Braun J (eds) InternationalesSymposionInterpraevent, Genoble, April 23-26. Klagenfurt, InternationaleForschungsgesellschaftInterpraevent, pp 1033–1043

Papathoma-Köhle M, Neuhauser B, Ratzinger K, Wenzel H, Dominey-Howes D (2007) Elements at risk as a framework for assessing vulnerability of communities to landslides. Nat Hazards Earth Syst Sci 7:765–779

Natural Hazards

Papathoma-Köhle M, Kappes M, Keiler M, Glade T (2011) Physical vulnerability assessment for Alpine hazards - state of the art and future needs. Nat Hazards 58:645–680
Papathoma-Köhle M, Zischg A, Fuchs S, Glade T, Keiler M (2015) Loss estimation for landslides in mountain areas—an integrated toolbox for vulnerability assessment and damage documentation. Environ Model Softw 63:156–169. https://doi.org/10.1016/j.envsoft.2014.10.003
Papathoma-Köhle M, Gems B, Sturm M, Fuchs S (2017) Matrices, curves and indicators: a review of approaches to assess physical vulnerability to debris flows. Earth Sci Rev 171:272–288. https://doi.org/10.1016/j.earscirev.2017.06.007
Pham BT, Pradhan B, Tien Bui D, Prakash I, Dholakia MB (2017) Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. CATENA 149:52–63. https://doi.org/10.1016/j.catena.2016.09.007
Prochaska AB, Santi PM, Higgins JD, Cannon SH (2008) A study of methods to estimate debris flow velocity. Landslides 5:431–444. https://doi.org/10.1007/s10346-008-0137-0
Puissant A, Van Den Eeckhaut M, Malet JP, Maquaire O (2013) Landslide consequence analysis: a region-scale indicator-based methodology. Landslides 11:843–858. https://doi.org/10.1007/s10346-013-0429-x
Ragozin AL, Tikhvinsky IO (2000) Landslide hazard, vulnerability and risk assessment. In: Bromhead E, Dixon N, Ibsen ML (eds) Proceedings of the 8th international symposium on landslides, Cardiff, pp 1257–1262
Rheinberger CM, Romang HE, Bründl M (2013) Proportional loss functions for debris flow events. Nat Hazards Earth Syst Sci 13:2147–2156
Rickenmann D (1999) Empirical relationships for debris flows. Nat Hazards 19(1):47–77
Sajwan KS, Sushil K (2016) A Geological Appraisal of Slope Instability in Upper Alaknanda Valley, Uttarakhand Himalaya, India. J Geol Geophys 5(5):1–7. https://doi.org/10.4172/2381-8719.1000258
Sarkar S, Kanungo DP, Sharma S (2015) Landslide hazard assessment in the upper Alaknanda valley of Indian Himalayas. Geomat Nat Hazards Risk 6(4):308–325
Silva M, Pereira S (2014) Assessment of physical vulnerability and potential losses of buildings due to shallow slides. Nat Hazards 72:1029–1050
Singh A, Pal S, Kanungo DP, Pareek N (2017) An overview of recent developments in landslide vulnerability assessment—presentation of a new conceptual framework. In: Mikos M, Tiwari B, Yin Y, Sassa K (eds) Advancing culture of living with landslides. WLF 2017, vol 2., pp 795–802. https://doi.org/10.1007/978-3-319-53498-5_91
Singh A, Kanungo DP, Pal S (2018) A modified approach for semi-quantitative estimation of physical vulnerability of buildings exposed to different landslide intensity scenarios. Georisk. https://doi.org/10.1080/17499518.2018.1501076
Thennavan E, Ganapathy GP, Sekaran SSC, Rajawat AS (2016) Use of GIS in assessing building vulnerability for landslide hazard in The Nilgiris, Western Ghats, India. Nat Hazards 82(2):1031–1050
Thouret JC, Ettinger S, Guitton M, Santoni O, Magill C, Martelli K, Zuccaro G, Revilla V, Charca JA, Arguedas A (2014) Assessing physical vulnerability in large cities exposed to flash floods and debris flows: the case of Arequipa (Peru). Nat Hazards 73:1771–1815
Totschnig R, Sedlacek W, Fuchs S (2011) A quantitative vulnerability function for fluvial sediment transport. Nat Hazards 58(2):681–703
Uzielli M, Nadim F, Lacasse S, Kaynia AM (2008) A conceptual framework for quantitative estimation of physical vulnerability to landslides. Eng Geol 102:251–256
Uzielli M, Catani F, Tofani V, Casagli N (2015) Risk analysis for the Ancona landslide—II: estimation of risk to buildings. Landslides 12:83–100. https://doi.org/10.1007/s10346-014-0477-x
Voogd H (1983) Multi criteria evaluation for urban and regional planning. Pion Limited, London, p 367
Walker G, Deeming H (2006) Functional and technical architectural design of a decision-support system for risk informed spatial planning. Deliverable 5.2, ARMONIA

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

International Journal of New Technology and Research (IJNTR)
ISSN:2454-4116, Volume-4, Issue-12, December 2018 Pages 24-26

Principal Component Analysis Of TF-IDF In Click Through Rate Prediction

Ankita Pal

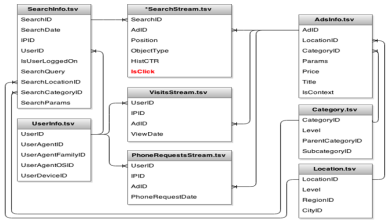
***Abstract*— This paper presents a model to predict the probability whether a user will click on a particular advertisement or not. The dataset used is that of Avito.ru provided as a part of the Kaggle competition- "Avito Contextual Ads Prediction". Here Principal Component Analysis on the Search and Query features is used, some extra count variables are made for integrate the categorical variables. Lastly logistic regression, SVM and gradient boosting algorithms are applied for classification into click or no-clicks.**

***Index Terms*— Avito Contextual Ads Prediction, Click Through Rate, Principal Component Analysis**

I. Introduction

Advertising is a major source of revenue for the companies which deals with search engines. These companies have a pay-per-click policy for charging the advertisers. As per the Pay-per-click policy, an advertiser has to pay for only those ads which have some clicks on them. They estimate the best possible location for placing ads so as to maximize revenue by using Click-through Rate (CTR). Hence, Click-through Rate prediction is very crucial to Sponsored Search.

In this paper, we have a training data set of 10.2GB and a test set of 531MB. For model formation we have taken small subsets of these data. The database has 8 tables as shown:



trainSearchStream.tsv and testSearchStream.tsv are two main data tables used in this paper of predictive model. They have a column ObjectType which shows which type of advertisement. Avito.ru supports 3 types of advertisements:

1. Regular: Free ads which go down with time
2. Highlighted: Advertisers pay a fixed amount for these ads. The ad stays on the top for a fixed amount of time then is shifted down.
3. Contextual: Advertisers pay per click. Therefore, it is necessary to predict the probability of click, so as to estimate the revenue from the ads.

In this paper we predict the probability of click of only the Contextual ads, which corresponds to ObjectType 3.

To reach our desired solution, we use the concept of Principal Component Analysis, a dimensionality reduction technique, on Term Frequency-Inverse Document Frequency weighted Document-Term matrix. Since, the Document –Term matrix is very large and very sparse, we need to reduce its dimensions to be able to use it in classification techniques. We also created some count variables of some categorical fields to be passed into the classification algorithm .We used various classification techniques and compare the results of various resultant models.

II. Methodology

A. *Data Pre-processing:*

1) *Data cleaning:* We have two text features –SearchQuery in SearchInfo and Title in AdsInfo. To integrate them into our predictive model, we need to prepare their Document Term Matrices (DTM) using the Term Frequency- Inverse Document Frequency weighting (TF-IDF). We make two corpuses, with each row of the two columns as documents of the respective corpuses. As a part of cleaning and pre-processing step we translate these corpuses from Russian to English and the remove punctuations, digits, stopwords, whitespaces and convert everything to lower case and also stem them.

2) *Feature engineering :*We make a DTM of the resulting 2 corpuses separately, using the TF-IDF weighting.

$$Wi,j = tfi,j \times log(\frac{N}{dfi})$$

Here, Wi,j =weight of i^{th} word in j^{th} document

tfi,j =number of occurrences of i^{th} word in j^{th} document

N= total number of documents

dfi=number if documents containing i^{th} word

We see the matrix is very large but very sparse. We need to reduce the dimensionality of the matrix , so we used Principal Component Analysis to reduce the number of dimensions and selected 2 each of the most relevant dimensions to represent the SearchQuery in SearchInfo and Title in AdsInfo.

In Principal Component Analysis,we decompose the DTM into Singular Value Decomposition (SVD) .

$$\begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}_{m \times n} = \begin{bmatrix} u_{11} & \cdots & u_{1r} \\ \vdots & \ddots & \vdots \\ u_{m1} & \cdots & u_{mr} \end{bmatrix}_{m \times r} \times \begin{bmatrix} s_{11} & \cdots & s_{1r} \\ \vdots & \ddots & \vdots \\ s_{r1} & \cdots & s_{rr} \end{bmatrix}_{r \times r} \times \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{r1} & \cdots & v_{rn} \end{bmatrix}^{T}_{r \times n}$$

Ankita Pal,Gurgaon, India

International Journal of New Technology and Research (IJNTR)
ISSN:2454-4116, Volume-4, Issue-12, December 2018 Pages 24-26

To get the principal components, we multiply:

$$B = V \times L^{\frac{1}{2}}$$
$$F = X \times B$$

Here F contains the principal components.
Another feature engineering step carried out was that of making count columns of LocationID, CategoryID and UserDeviceID to be passed into the classification algorithm.

B. *Classification Algorithms:*

1) *Logistic Regression:* It is one of the mostly widely used classification methods in CTR prediction. Binary Logistic Regression is a special type of regression where binary response variable is related to a set of independent variables, which can be discrete and/or continuous. In logistic regression the outcome is be 0 or 1. We apply the sigmoid function to the results of the linear regression model to get the dichotomous outcome.

$$f(x) = \frac{1}{1 + e^{\theta^T x}}$$

Here: f(x) =outcome variable
θ^T=matrix of weights
x =input vector

2) *Support Vector Machine (SVM)*: Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used classification as well as regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well.

3) *Gradient Boosting:* Gradient Boosting Machines(GBMs) are the learning procedure that consecutively fits new models to provide a more accurate estimate of the response variable. The idea behind this algorithm is to construct the new weak learners to be maximally correlated with the negative gradient of the loss function, associated with the whole ensemble. The loss functions applied can be arbitrary, but to give a better intuition, if the error function is the classic squared-error loss, the learning procedure would result in consecutive error-fitting.

C. *Evaluation metrics:*

The evaluation metric used was the log loss function. It is a classification loss function

$$\log loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij} log(p_{ij})$$

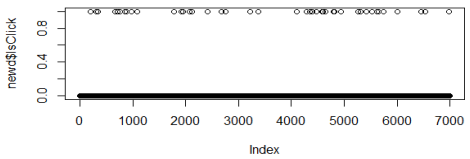
Here: N=number of samples
y_ij=is a binary indicator of whether or not label j is the correct classification for instance i
p_ij =is the model probability of assigning label j
When there are only 2 classes i.e. it is a Binary Classification then:

$$\log \text{loss} = -\frac{1}{N}\sum_{i=1}^{N} y_i log(p_i) + (1 - y_i) log(1 - p_i)$$

Here: N=number of samples
y_i =is a binary indicator of whether ad is clicked or not
p_i =is the model probability of the advertisement being clicked

Log-loss measures the accuracy of a classifier. It is used when the model outputs a probability for each class, rather than just the most likely class. Log loss measures the uncertainty of the probabilities of your model by comparing them to the true labels.

III. Results



Plot of Clicks in first 7000 records with ObjectType 3

From figure we notice that the number of no-clicks is far more than clicks. Thus, we expect the click through rate to be very small numerically. The probability of click lies between 0 and 1 for all records in all models. We are more interested in how much close the probability is close to actual: 0 for not being clicked, 1 for being clicked.

We compare the log loss we get from the 3 models we made using the 3 classification models:

S.No	CLASSIFICATION ALGORITHM	LOG LOSS	Rank
With the text features reduced by PCA of TF-IDF weighted matrix, without count features.			
1	Logistic Regression	0.0331649	2
2	Support Vector Machine(SVM)	0.0871754	3
3	Gradient Boosting	0.0322150	1
With the text features reduced by PCA of TF-IDF weighted matrix, with count features.			
1	Logistic Regression	0.0326645	2
2	Support Vector Machine(SVM)	0.1252218	3
3	Gradient Boosting	0.0322914	1

IV. Conclusion

We observed the performance of various classification models:

- With the text features reduced by PCA of TF-IDF weighted matrix, without count features.
- With the text features reduced by PCA of TF-IDF weighted matrix, with count features.

We conclude that the best model is with Gradient Boosting, followed by Logistic Regression.

We notice that SVM is not an appropriate classifier.

The reasons for the good log loss scores for Gradient Boosting is that it is an ensemble method and increases the accuracy by learning from weak classifiers, and does not need feature engineering. By feature engineering we further tune up the model for better predictions.

The main challenge was the large dataset provided as a part of the competition, which could not be handled very efficiently, given the restricted resources. So we took a subset of 10000 records for the project. The DTM for larger records could not be stored in memory. Another challenge was that of conversion of Russian data to English, which was done using Google Translate.

References

[1] Steffen Rendle. Scaling factorization machines to relational data. In Proceedings of the VLDB Endowment, volume 6, pages 337–348. VLDB Endowment, 2013.
[2] H Brendan McMahan, Gary Holt, David Sculley, Michael Young, Dietmar Ebner, Julian Grady,Lan Nie, Todd Phillips, Eugene Davydov, Daniel Golovin, et al. Ad click prediction: a view from the trenches. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 1222–1230. ACM, 2013.
[3] Zhipeng Fang, Kun Yue, Jixian Zhang, Dehai Zhang, and Weiyi Liu. Predicting click-through rates of new advertisements based on the bayesian network. Mathematical Problems in Engineering,2014, 2014.
[4] Norwegian Computing Center, P.B. 335 Blindern, N 0314 Oslo 3 (Norway) and Research Group for Chemometrics, Institute of Chemistry, Umed University, S 901 87 Umeci (Sweden). Principal Component Analysis
[5] AlexeyNatekin and AloisKnoll Gradient boosting machines a tutorial.2013 available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3885826/



Ankita Pal B.Tech in Mathematics and Computing from Delhi Technological University(2013-2017). General Secretary of Society of Industrial and Applied Mathematics for the period of 2016-2017. Currently working in Axtria India Pvt. Ltd.

ISSN 2007-9737

Semi Supervised Graph Based Keyword Extraction Using Lexical Chains and Centrality Measures

Ayush Aggarwal[1], Chhavi Sharma[1], Minni Jain[1], Amita Jain[2]

[1] Delhi Technological University, Delhi, India

[2] Ambedkar Institute of Advanced Communication Technologies and Research, Delhi, India

{aggarwal96ayush, sharma.chhavi96}@gmail.com
minnijain@dtu.ac.in, amita_jain_17@yahoo.com

Abstract. This paper presents keyword extraction using lexical chains and graph centrality measures, derived from the semantic similarity of the words by analysis of the graphical network created using WordNet. The hypothesis is presented using a small-world approach where every paragraph in a document is constrained to a local point, while the document in all is centered on a global concept. Creating lexical chains for each paragraph and combining the best via scoring methods and graph based algorithms, we present parallels to baseline system to extract the keywords from the document.

Keywords. Graph centrality, keyword extraction, lexical chains, semantic similarity, small world approach, WordNet.

1 Introduction

Keyword extraction is used to describe the basic subject of a document by identification of terms that best describe it [2, 10]. This research focuses on an algorithm for automatic extraction of these keywords and thus provides with the document's summary. Unlike text summarization, which provides certain important extracts of the document, keyword extraction also extends the functionality by helping in the indexing of document for search engines or text categorization by providing singular words that best represent the document. In the present day scenario, where there is a lot of data available, it becomes seemingly impossible to sift through the entire collection manually, with the aim of finding relevant information whereas keywords allow us to identify the main point of the author.

Unfortunately, in many cases we do not have labeled text pieces with their keywords and thus keyword extraction helps automate the task. In this paper, we focus our work on *keywords* rather than *key-phrases* as key phrases are composed of more than one word while we are extracting only the keywords.

Existing Automatic keyword extraction tools provide generally the words with highest occurring frequency. Moreover they do not correlate with various synsets (groups of synonymous English words) of the same word, thus resulting in multiple keywords of the same root-word or failing in classifying a word as a keyword due to its occurrence multiple times, but as different synsets of the same root-word [16]. These problems can be removed by using WordNet corpus, where there is an existing database of synsets, along with various relations to other words such as hyponymy, hyperonymy and hypernymy.

Lexical chains are sets of semantically related words [17, 21]. The keywords of the document would be frequently occurring as synsets and hence, would be contained in the lexical chains of the paragraph. A chain can be representative of a small portion portion of the document. Thus various chains are calculated across the document and are scored to identify their relevancy. The chains with higher scores are considered to be representatives of the document.

The similarity between keyword extraction and text summarization roots from the fact that both the NLP techniques take parts of documents to represent the summary. Lexical chains have been shown to be used for text summarization [1, 4] in the past. Recent researches have started to explore the use of lexical chains to extract keywords from a document. Basing the hypothesis on these, we propose an algorithm for keyword extraction using WordNet ontology.

ISSN 2007-9737

2 Related Work

Various methods of keyword extraction have been tried extensively, especially for the purposes of indexing and improving web searches of the documents. Keyword extraction can be classified into supervised, semi-supervised and unsupervised approaches.

Supervised methods work with a hand annotated data-set of documents and keywords and use domain specific knowledge to classify them in two classes (or binary classification): *keyword or not*. Two standard supervised learning systems in the field of keyword extraction are KEA [23] and GenEX [18]. Both of these approaches are based upon the frequency as well as the location of term in a document while classifying. While GenEx uses a C4.5 Decision tree to classify, KEA uses Naive Bayes for learning and classification. KEA has been further improved over the years using statistical association between key phrases [19] or by using semantic information from a domain specific thesaurus as in KEA [14].

The issues with supervised keyword extraction methods are a need for hand annotated keywords for training purposes and a bias towards the domain specific knowledge base they are trained upon. That is why we now aim at semi-supervised approaches. Graph Based approaches provide a sophisticated way to extract keywords using the structure of the graph created using source statistics.

KeyGraph is an algorithm for automatic indexing by a co-occurrence graph constructed from metaphors [9]. It is based upon segmenting of a graph, representing co-occurrences of the document as clusters. Each cluster represents basically the Small World structure of the document.While the small world structure helped to model and connect different meanings of the document into a global meaning, it was a content sensitive and domain independent algorithm.

SemanticRank uses a graph based technique exploiting the semantic relatedness of the document using knowledge-based measures of WordNet and Wikipedia and graph centrality measures of PageRank [15] and Hits [7]. The success of the above Graph based approaches using WordNet as their knowledge base by performing at par with the supervised learning standard systems led us to incorporate the same in our work.

The motivation behind using lexical chains for Keyword Extraction is behind the fact that lexical chains have been used for text summarization [1] and that there is a great sense of similarity between text summarization and keyword extraction. Prior work on using lexical chains for keyword extraction [3] had shown good preliminary results, which shows that this can be further looked into. But this method used a supervised learning approach to create lexical chains and hence does not solve the problem of non-availability of hand tagged data-set and is highly biased to multiple occurrence of words belonging to same synset or having the same root.

In the proposed work, we combine the graph based semi supervised approach with the benefits of using lexical chains to propose an algorithm that performs much better than both the approaches if handled singularly. It not only makes the algorithm domain independent and free of needing prior knowledge of the domain to extract information but the small world approach of treating each paragraph as a building base of the whole document helps to distribute the global meaning in the proposed algorithm thus reducing bias of the system.

Computación y Sistemas, Vol. 22, No. 4, 2018, pp. 1307–1315
doi: 10.13053/CyS-22-4-3077

3 The Lexicon

We use a database of English words and their relations: WordNet[1] [11]. It uses the concept of "cognitive synonyms" or synsets, and interlinks them on the basis of lexical relations. It is a free linguistic tool which acts as a dictionary or thesaurus, providing the user with synonym sets, word definitions and their examples.

3.1 Structure

WordNet connects words using various relationships. The main relation among these is *synonymy* on the basis of which we derive various synsets : the basic unit of WordNet relational identity [11]. Each synset, thus derived from the relation of synonym, in turn, is linked to various other synsets by the relations described below.

3.2 Relations

One of the most important relations among the semantically related synsets is the relation of super-subordination called *hyponymy*. While hyponymy connects the broader sense of the word with a specific or a part of the sense, *hypernymy* does the opposite and connects the synsets using a is-a relation [5]. The following Figure 1 shows a semantic network constructed using the relations of *hyponymy* and *hypernymy* of the word dog.

4 Lexical Chains

A lexical chain can be defined as a sequence of related words that denote the semantic context of the piece of text [12]. Hence, determining these chains helps to identify the main topics of a document. They have been previously explored for information retrieval and related areas [22]. In the proposed work, we will be using WordNet to build lexical chain as done by Stairmand [17] as lexical chains require the use of an ontology or a database which has predefined chains of semantically similar words.

Generally the procedure for creating a lexical chain is as follows as postulated by Stairmand [17]:

1. Select a set of candidate words.
2. Determine a suitable chain, calculated from semantic relatedness among members of the chain for each selected word.
3. If a chain exists, add the word and update the chain else create a new chain to fit the word.

The semantic similarity of the proposed algorithm is based on a lexical similarity measure created by Wu and Palmer [24]. The measure examines which lexical chain provides the maximum relevance to the given word and whether it passes a threshold value.

4.1 Performance Function

After creating a lexical chain, we need to score them to determine the best lexical chains to determine the words to be fed into our graph based system. The lexical chains were scored using various methods, and we selected the one with better precision as shown in the next sections.

Method I:
This method builds upon the methods described by LexSum[**?**]. In this scoring method, the score of a chain is the summation of score of all the members in the chain divided by the numbers of members. The score of each member in the chain is a combination of frequency of the word in the paragraph and its similarity index with the members of the chain:

$$chain_score(C) = \frac{\sum rel_i(w) * freq(w)}{length(C)}, \quad (1)$$

where $rel_i(w)$ denotes the similarity index of member w with i which is the representative of that chain: the member with maximum frequency in that chain. Now the threshold value of a paragraph is calculated:

$$threshold(P) = \mu(c) * \sigma(c) : \forall c \in C, \quad (2)$$

and for every chain whose score is greater than the aforementioned threshold, we add all the member's representative words with similarity score greater than threshold to the final list of words on which we apply the proposed graph based algorithm.

Method II:
We propose another method for scoring of lexical chain. This method of scoring lexical chain is dependent not on the frequency of the word in the paragraph, as common words with broad meaning appear more, but on the semantic distance between the words of each chain. For each member of the lexical chain, we maintain a list of hypernyms and hyponyms, and thus find the semantic distance of a lexical chain by iterating through the lexical chain and scoring a hypernym/hyponym match with *0.5* and an identical/ synonym match with *1*.

Let the set of lexical chains be denoted by $L = \{L_1, L_2, ..., L_n\}$ and for every lexical chain we say $L_i = \{w_1, w_2, ..., w_n\}$ be the set of connected words where every word will has a score of c_j, then:

$$chain_score(L_i) = \frac{\sum c_j}{length(L_i)} \forall c_j \in L_i. \quad (3)$$

We calculate the threshold of all the chains of the paragraph as:

$$threshold(P) = \mu(L_i) + 2 * \sigma(L_i) : \forall L_i \in L, \quad (4)$$

as taken by Braizaley [1] and for every chain whose score crosses the threshold , we select the representative of the said chain where representative refers to the member with maximum member score in the chain.

5 Proposed Scheme

The algorithm that we propose incorporates the features from a lexical chain based algorithm, which makes the algorithm identify the emphasized sense of the document by creating lexical chains

[1] WordNet is available from http://wordnet.princeton.edu

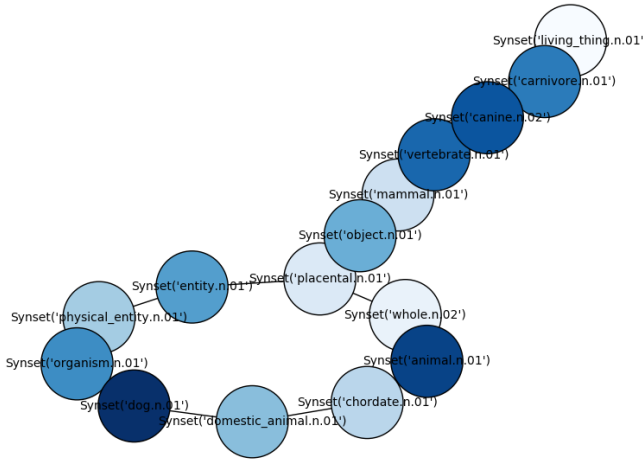


Fig. 1. Graph extracted from WordNet using *synsets* of the word *dog*

for every paragraph and using the best chains to feed it to a graph based WSD algorithm as suggested by Navigali and Lapata [13]. The advantage of using the graph approach on top of a lexical chain based segregation is that it removes the shortcomings of the supervised learning algorithm and makes it domain independent and not in requirement of a pre-fed training data to adapt. The use of WordNet assures us that the algorithm can be extended to any use-case. The keyword extraction algorithm uses a small world approach to ensure the central meaning of the document is identified.

The algorithm parses every paragraph of the document and first extracts all the nouns (proper nouns and compound nouns) from the paragraph. Now using these lists of nouns, lexical chains for a particular paragraph are built as follows:

— Calculate the similarity of the current word with all the previously created lexical chains.

— If the maximum similarity value is greater than a threshold value, then add this word to the lexical chain else create a new lexical chain with the word as its first entry.

After creating the lexical chains for each paragraph they are scored and then strength, average score and standard deviation for these lexical chains are calculated separately. Using these values, we calculate the cut-off score for lexical chains in a particular paragraph and determine the *best chains* i.e. all the lexical chains which have a score greater than this cut-off score.

We determine these best chains for all paragraphs and add all the members of these chains to a word list which will be the probable list of keywords. We feed this word list to the graph algorithm that finds semantic similarity between members of different lexical chains.

The algorithm to construct a graph is built upon the method used in word sense disambiguation [13]. The method has been proposed to be used in more language processing tasks [6] and the idea behind this algorithm is to find most important node using network analysis. We build a graph $G = (V, E)$ using the word senses as the nodes and the semantic relation between them as the un-directed edges. We then select a sense u ϵ senses and execute the following algorithm on it:

— Generate sets of two relations, hyponymy and hypernymy u.

— Perform a Depth First Search for all the words v in those sets. Every time we encounter a node belong to destination along the path, we plot it on the graph.

The algorithm does a DFS from the last vertex processed and uses a recursive approach to explore the un-visited vertices. It involves backtracking to a previous state if all the edges connecting a node have already been visited before. This process continues till we reach a state where no further discovery can be made. Centrality measures help to calculate the degree of relevance of a vertex v in a graph G. It also calculates the influence of the vertex in its network. These measures are pivotal for the algorithm as they analyze the connectivity of the probable keywords with all the other words from the set to identify the most central word as the extracted keyword from the document.

5.1 Proposed Algorithm

Let a document D be a set of paragraphs $P = \{P_1, P_2, ..., P_n\}$, then for every graph $P_i \in D$:

Step I: Extract all nouns using POS and create lexical chains from the set of nouns n.
Step II: To create a lexical chain, pick a word from the set of nouns and calculate wup_similarity with the existing chains. Add it to one of the existing chains or start a new chain.
Step III: For all the lexical chains $L = \{L_1, L_2, ..., L_k\}$, calculate the chain score by using equation 1 OR 3 and create a word list W which are a set of potential keywords.
Use the word list W to create a semantic network using hyponymy, hypernymy and synonymy relation as follows:
Step IV: For every $w_i \in W$, add the word to the graph G.
Step V: For every$w_i \in W$, calculate the hyponyms and hypernyms and synonyms and do a DFS with

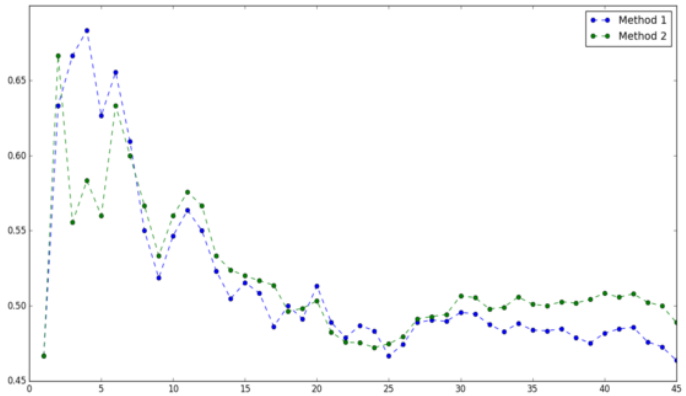


Fig. 2. Scoring Method 1 and Method 2

all the discovered words. If a match is found, a relationship is said to exist between two words.
Step VI: Explore the edges from most recently discovered word w_j till the point there are no new edges to be discovered.
Step VII: After the words are processed, analyze the network using the following centrality methods. If w_i, w_j denotes the nodes of a graph and E the set of edges and W the set of nodes then:

$$Degree(w_i) = |(w_i, w_j) \in E : w_i, w_j \in W|, \quad (5)$$

$$Betweenness(w_i) = \sum \frac{\sigma_{i,j}(w_i)}{\sigma_{i,j}}, \quad (6)$$

$$PageRank(w_i) = \frac{1-k}{|W|} + k \sum_{w_i, w_j \in E} \frac{PR(w_j)}{outdegree(w_j)}, \quad (7)$$

$$Closeness(w_i) = \frac{\sum_{w_j \in W: w_i \neq w_j} \frac{1}{d(x,y)}}{|S| - 1}, \quad (8)$$

$$HITS_A(w_i) = \sum_{w_j \in In(w_i)} HITS_H(w_j), \quad (9a)$$

$$HITS_H(w_i) = \sum_{w_j \in Out(w_i)} HITS_A(w_j), \quad (9b)$$

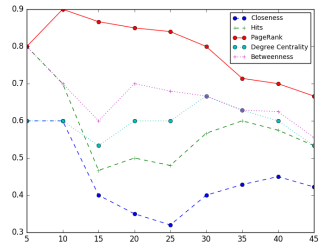
and evaluate the top keywords from every centrality to propose result.

6 Simulation Results

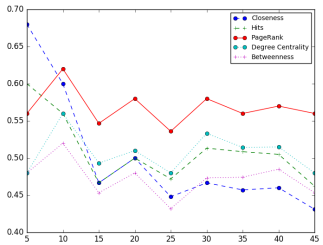
We present the research as an improvement for already existing work by Ercan and Cicekli (2007) [3]. Since it was a supervised learning algorithm, we use another dataset and their baseline results to present our comparison. We use 500N-KPCrowd [8] for testing the hypothesis. The data-set is a corpora of news articles from 10 different topics with hand annotated keywords associated with every news report. We use 45 articles to run the algorithm and aggregate results to present a direct comparison with the baseline results. It also has a list of hand annotated list of keywords by 20 Amazon workers.

We also use two different scoring methods and present a parallel between the two to consider for our algorithm as proposed above. From following figure 2 we notice that as the number of articles in the data-set increases, the accuracy of Method 2 supersedes that of the first method.

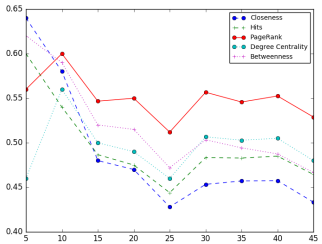
This is in accordance with our hypothesis that using small world approach thus makes the algorithm less biased towards word with more frequency in the text. This method also provides better results because it also takes into



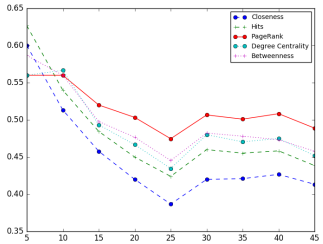
(a) 1 keyword



(b) 5 keywords



(c) 10 keywords



(d) 15 keywords

Fig. 3. Comparing precision score for different centrality measures for varying amount of keywords extracted

consideration the strength of the relation between two words and scores it accordingly.

For comparison with baseline results, we run our algorithm to extract various number of keywords on the same documents and compare the precision score thus achieved for various centrality measures. The results are thus plotted on the graph and we compare the performance of received results.

As we depict in the figure 3, PageRank centrality method surpasses other centrality measures as compared to the other centrality methods. This is in line with previous works [20] which also say that PageRank out of all other centrality measures can be be used to extract keywords from graph based network.

Table 1. Table for results observed

.

	Baseline Results	Our Algorithm
1 keyword	64.0%	66.7%
5 keywords	45.0%	56.0%
10 keywords	30.0 %	52.8%
15 keywords	26.0%	48.8%

We also compare our results with the baseline results as proposed by Ercan and observe that the algorithm surpasses the baseline results with proving to be a better improvement for extracting higher number of keywords. This is depicted in the Table 1.

Using graph based centrality method after using lexical chains improves the already existing

baseline systems with considerate improvement when more keywords are being extracted for the document as proposed.

7 Conclusion and Future Work

We developed and discussed a graph based method, which supersedes the accuracy of pre-existing supervised learning algorithms. By incorporating the advantages of a network analysis approach using centrality methods and lexical chains, we were able to develop an algorithm, which is extendable to any domain and is less biased towards the type or frequency of word used in the document. However, this algorithm currently revolves around extraction of keywords from the document. It fails, when there is a set of words or *key-phrases*. Hence, we would like to expand to extraction of key phrases from a document.

The algorithm is unable to include words, which are not defined in the corpus and for this we plan to look into further knowledge bases such as Wikipedia.

References

1. **Barzilay, R. & Elhadad, M. (1997).** Using lexical chains for text summarization. *In Proceedings of the ACL Workshop on Intelligent Scalable Text Summarization*, pp. 10–17.
2. **Beliga, S., Meštrović, A., & Martincic-Ipsic, S. (2015).** An overview of graph-based keyword extraction methods and approaches. *Journal of Information and Organizational Sciences*, Vol. 39, pp. 1–20.
3. **Ercan, G. & Cicekli, I. (2007).** Using lexical chains for keyword extraction. *Inf. Process. Manage.*, Vol. 43, No. 6, pp. 1705–1714.
4. **Erkan, G. & Radev, D. R. (2004).** Lexrank: Graph-based lexical centrality as salience in text summarization. *J. Artif. Int. Res.*, Vol. 22, No. 1, pp. 457–479.
5. **Franzoni, V., Li, Y., Leung, C. H. C., & Milani, A. (2017).** Semantic evolutionary concept distances for effective information retrieval in query expansion. *CoRR*, Vol. abs/1701.05311.
6. **Jain, A. & Lobiyal, D. K. (2015).** Fuzzy hindi wordnet and word sense disambiguation using fuzzy graph connectivity measures. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, Vol. 15, No. 2, pp. 8:1–8:31.
7. **Kleinberg, J. M. (1999).** Authoritative sources in a hyperlinked environment. *J. ACM*, Vol. 46, No. 5, pp. 604–632.
8. **Marujo, L., Gershman, A., Carbonell, J. G., Frederking, R. E., & Neto, J. P. (2013).** Supervised topical key phrase extraction of news stories using crowdsourcing, light filtering and co-reference normalization. *CoRR*, Vol. abs/1306.4886.
9. **Medelyan, O. & Witten, I. H. (2006).** Thesaurus based automatic keyphrase indexing. *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '06, ACM, New York, NY, USA, pp. 296–297.
10. **Mihalcea, R. & Tarau, P. (2004).** Textrank: Bringing order into text. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*.
11. **Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990).** WordNet: An on-line lexical database. *International Journal of Lexicography*, Vol. 3, pp. 235–244.
12. **Morris, J. & Hirst, G. (1991).** Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Comput. Linguist.*, Vol. 17, No. 1, pp. 21–48.
13. **Navigli, R. & Lapata, M. (2010).** An experimental study of graph connectivity for unsupervised word sense disambiguation. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 32, No. 4, pp. 678–692.
14. **Ohsawa, Y., Benson, N. E., & Yachida, M. (1998).** Keygraph: Automatic indexing by co-occurrence graph based on building construction metaphor. *Proceedings of the Advances in Digital Libraries Conference*, ADL '98, IEEE Computer Society, Washington, DC, USA, pp. 12–.
15. **Page, L., Brin, S., Motwani, R., & Winograd, T. (1998).** *The PageRank Citation Ranking: Bringing Order to the Web.*
16. **Ramos, J. (2003).** *Using TF-IDF to Determine Word Relevance in Document Queries.*
17. **Stairmand, M. & Engineering, L. (1996).** *A computational analysis of lexical cohesion with applications in information retrieval.* UMIST.

18. Turney, P. D. (2000). Learning algorithms for keyphrase extraction. *Inf. Retr.*, Vol. 2, No. 4, pp. 303–336.

19. Turney, P. D. (2003). Coherent keyphrase extraction via web mining. *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, IJCAI'03, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 434–439.

20. Wang, J., Liu, J., & Wang, C. (2007). Keyword extraction based on pagerank. **Zhou, Z.-H., Li, H., & Yang, Q.**, editors, *Advances in Knowledge Discovery and Data Mining*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 857–864.

21. Wei, T., Lu, Y., Chang, H., Zhou, Q., & Bao, X. (2015). A semantic approach for text clustering using wordnet and lexical chains. *Expert Syst. Appl.*, Vol. 42, No. 4, pp. 2264–2275.

22. Wei, T., Lu, Y., Chang, H., Zhou, Q., & Bao, X. (2015). A semantic approach for text clustering using wordnet and lexical chains. *Expert Syst. Appl.*, Vol. 42, No. 4, pp. 2264–2275.

23. Witten, I. H., Paynter, G. W., Frank, E., Gutwin, C., & Nevill-Manning, C. G. (1999). Kea: Practical automatic keyphrase extraction. *Proceedings of the Fourth ACM Conference on Digital Libraries*, DL '99, ACM, New York, NY, USA, pp. 254–255.

24. Wu, Z. & Palmer, M. (1994). Verb semantics and lexical selection. *CoRR*, Vol. abs/cmp-lg/9406033.

Article received on 19/12/2017; accepted on 15/02/2018.
Corresponding author is Ayush Aggarwal.

arXiv:1812.11606v1 [cs.CV] 30 Dec 2018

Solar Potential Analysis of Rooftops Using Satellite Imagery

Akash Kumar and S. Indu

Delhi Technological University, New Delhi
{akash_bt2k15, s.indu}@dtu.ac.in

Abstract. Solar energy is one of the most important sources of renewable energy and the cleanest form of energy. In India, where solar energy could produce power around trillion kilowatt hours in a year, we are producing only power of around in gigawatts only. Many people are not aware of the solar potential of their rooftop and hence they always think that installing solar panels is very much expensive. In this work, we introduce an approach through which we can generate a report remotely that provides the amount of solar potential of a building using only its latitude and longitude. We further evaluated various types of rooftops to make our solution more robust. We also provide an approximate area of rooftop that can be used for solar panels placement and a visual analysis of how solar panels can be placed to maximize the output of solar power at a location.

Keywords: Rooftop Detection · Solar Panels · Adaptive Canny Edge · Gabor Filter · Segmentation.

1 Introduction

Solar power is the process of conversion of sunlight energy to electrical energy. Photovoltaic cells present on the solar panels convert the sunlight energy to Direct current and then an inverter converts it to Alternative current source that is used in a household. Solar energy has several benefits ranging from habitat conservation to combat climate change. It is also the cheapest and reliable form of energy. Once installed, the solar panels can be sued up to 20-25 years with minimal maintenance. However, in India people hesitate to solar panels as they have to go through a lot of processes. The company will send their personnel and they will take each measurement and then build a report them. Sometimes, they have to take multiple visits too. It takes a lot amount of time and wastage of money too. It is also impossible for an industry to individually analyze an area by surveying the solar potential of every house. Technology like Image Processing and Computer Vision plays a pivotal role to overcome this challenge.

Many works have been done in foreign countries to automate the solar power analysis of an area. In the USA, Project Sunroof has been launched by Google to overcome this problem. They have used 3D mapping technology alongwith a high-quality satellite map that is available for the USA. Still, their technology

is limited to forty-two cities of USA only. Therefore, there is a need to create a more generalized solution that can adapt well to many situations.

In this paper, we employ several algorithms to address the problem of rooftop detection. Our main contributions are:

1. We proposed an approach to segment out optimal rooftop area for solar panel placement in India where the image quality of satellite image is very low.
2. We considered several types of rooftop to learn the intra-class variations and to evaluate our method. Our final output provides an approximate analysis of area that can be used for solar panel placement. Our solution also layouts at which angles solar panels should be placed for maximum solar output.

2 Related Work

There are several recent papers that have addressed the problems of identification of building rooftops [1], [2], [3], [5], most of them are based on 3D reconstruction, which uses Fast Graph search [1], GIS maps[2] and other topology based automatic modeling. [3] identified the rooftops using a single remote sensing image compared to a stereo-based approach. Markov Random Field and projective geometry were used to detect rooftops accurately.

Work has been done using LiDAR technology using Object-based [4] method to localize rooftops over an area. LiDAR technology automatically extracted building footprints and segment out the rooftop planes but this technology is very expensive and everyone could not afford to do that. Using Spectral Graph theory [22], automatic building detection has been done but that's also done over the USA dataset. Moreover, the houses in the USA have slanted rooftops whereas in India the rooftops are planes. Many works have been done on building detection and its 3D modeling in the USA but our method automatically detects the optimal area in case of Indian rooftops. The above methods failed miserably in case of planar rooftops. We explored the area of obstacle identification in the rooftop area using Satellite imagery and thus marks out the area available for solar power generation.

3 Problem Formulation

Solar panels placement in case of India is very different as compared to the USA. In the USA, Google Satellite Maps can be zoomed up to level 21 and they also have a 3D mapping that helps to localize rooftops as well as obstacles(e.g. Water tanks, AC Condenser units, etc.) whereas in India the maps can be zoomed up to level 20 only with no 3D mapping. Example of Google Satellite map at highest zoom level in the USA and in India is shown in Fig.1 below. From the figure, we can see the quality of maps in the USA as compared to India. While we can visually mark out the optimal area for solar panel placement in the USA, in India we can't even crop out the exact rooftop visually. Therefore, our solution focusses

on detection of the usable rooftop area of a building as well as to maximize the solar power generation by orienting solar panels at a perfect angle.

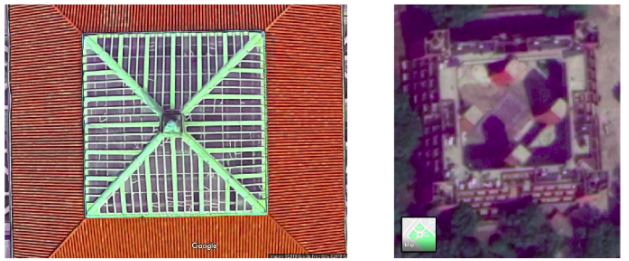


Fig. 1. Left Image depicts USA rooftop at Zoom Level 21 and right image depicts Indian building rooftop at Zoom Level 20.

4 Dataset

The manually collected Indian aerial rooftop dataset consists of all types of variants. We considered all the possibilities of the types of rooftop obstacles and also different rooftop sizes. Currently, we have tested our algorithm on 50 different types of rooftops. Overview of the dataset is shown in the diagram Fig.2.



Fig. 2. Dataset Overview

5 Proposed Approaches

In this section, we discuss the proposed framework that is used for optimal rooftop detection from the above rooftop images. The devised architecture used to solve the problem is described as follows:

5.1 Aerial Rooftop Detection

Watershed Segmentation This algorithm is mainly applied when we want to segment out objects-of-interest that are close to each other, that is their edges touch. It is based on the marker-based algorithm. It labels the region sure as foreground and background and the region not sure of as zero. Then, it updates the labels as segmentation procedure proceeds. It is an interactive segmentation process. It was used to count the number of buildings and segment out rooftop locations from the map. The algorithm was giving many false positives as depicted in Fig. 3.

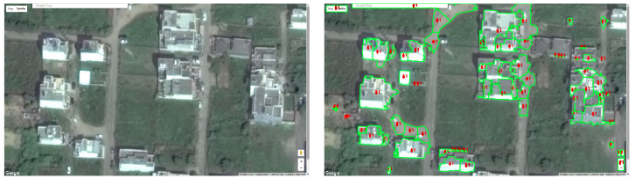


Fig. 3. Watershed Segmentation algorithm on Google Satellite Image on the left. On the right side, the output shows detection of buildings.

Gabor Filter It analyzes the region of interest and detects whether there is any specific frequency content in a particular direction is present or not.[6] This procedure of segmentation is based on Gaussian distribution technique. Gaussian Mixture Models(GMM) assumes all data points are a combination of Gaussian distribution functions. Gabor filter enhances one region relative to another depending on the frequency and theta values. After applying this filter, two GMMs were fit in the image histogram to separate out foreground and background. The output after applying two models is shown in Fig. 4.

Active Contours This algorithm is also known as *snakes algorithm* is used to find object boundaries within an image. [7] [8] [9] It is an iterative flow of gradient vector from the point of initiation in all directions. It has a resistance of bending

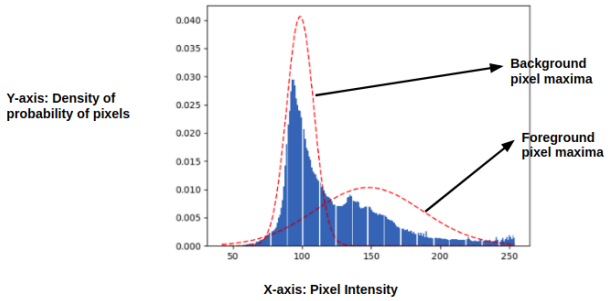


Fig. 4. Two Gaussian Mixture models were fixed over the satellite image to separate the foreground and background.

and stretching towards object boundaries. Hence, in our case it successfully detects the obstacles present inside the roof as well as mark the boundaries too. The result of Active Contour applied on edge sharpened image is shown in Fig. 5.

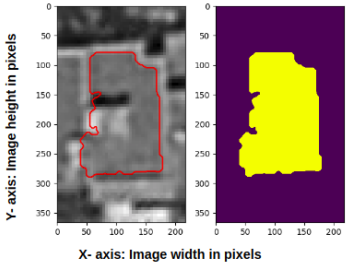


Fig. 5. Active Contours rooftop segmentation.

Adaptive Canny Edge Detection Based on the comparative analysis of edge detection algorithms, Canny edge detection algorithm outperforms others. Now, adaptive canny edge algorithm models itself based on the image.[10] The hys-

teresis threshold, in this case, is adaptive relevant to the image. The upper and lower threshold is calculated by using mean and variance of an image intensity. The output of the Canny Edge Detection algorithm also contains many false positives and it was also very haphazard as shown in Fig.6.

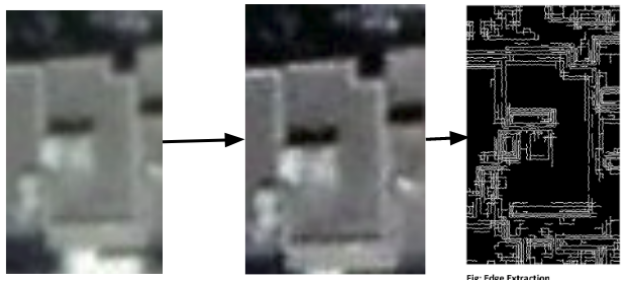


Fig. 6. Adaptive Canny Edge Detection results. The leftmost image is the original image. The center image is the bilateral sharpened image and the rightmost image is the output of the Edge Detection algorithm.

5.2 Polygon Shape Approximation

The shape of the polygon that we get from Active Contours as output is very much distorted. We devised the following procedure to obtain a perfect polygon.

1) Hough Transform It is used to detect any shape that can be represented in a mathematical form.[11] [12] It was used to analyze the shape of a rooftop. Using K-Means clustering, the number of Hough line Transform were reduced to 4 to 6 to outline the rooftop shape alongwith objects present inside the rooftop.

2) Region-based Polygon filling After applying Hough Transform in combination with K-Means clustering, the rooftop area was divided into different regions. Checking the intensity of different patches, the area was marked as a rooftop area or not. If the mean intensity is greater than the threshold value, then that region is marked as a rooftop area and it is filled with black color on a different white image. The approximate shape of a rooftop after applying Region-based filling is shown in Fig.7.

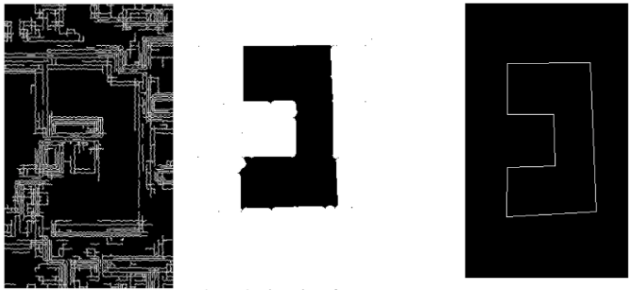


Fig. 7. Polygon Shape Approximation on Canny Edge detection output using K-means Clustering Algorithm.

5.3 Optimal Rooftop Area

Using single feature characteristic of an image is not much help to carve out the optimal rooftop area. It always leads to many false positives and it was not able to detect obstacles present inside the rooftops. Therefore, we employed two features at the same time to get the exact rooftop area. The two methods which we used to get the shape of the rooftop is as follows:

1) Harris Corners & Adaptive Canny The locations at which Harris corners and Canny Edge detection feature points overlap, those features were selected. The problems occurred with Harris corners as they can't be accessed in a localized manner. Therefore, to draw an exact polygon shape was not possible.

2) Adaptive Canny & Contours The output of Contours and Canny Edge features are in a clockwise manner. We applied Contours on two images. One is Bilateral Edge sharpened and the other is the Canny Edge map. Contours in Bilateral Edge sharpened image provides us the rooftop boundaries and in canny edge map, it gives obstacle boundaries present on the rooftop. Combining both the results and plotting it on a white patch gives the exact rooftop optimal area for solar panel placement.

6 Experiments and Results Analysis

Edge Detection Algorithms We evaluated various types of edge detection algorithms to detect rooftop boundaries.[14] [15] [16] Canny outperforms other algorithms such as Laplacian of Gaussian, Sobel, Robert and Prewit.

Solar Panel Placement We considered solar panels patches of 5x1, 4x1, 3x1. We also took into account the orientation angle required for maximum solar power output at a certain longitude and latitude. After placing solar panels horizontally, the main challenge was to rotate the panels with a certain angle, because, it will go out of the rooftop. For rotation of patches, I took a region of interest then rotated it by the amount of angle required. Then, we iterated along the breadth and length and evaluated the mean intensity of the region of interest using line iterator. If the mean intensity is 255, then the area is available for solar panel placement. The experimental results on the rooftops dataset are shown in the Fig. 8.

7 Conclusion and Future Work

The paper presented several approaches on which comprehensive analysis was done to segment out optimal rooftop area and placement of solar panels too. In India, due to the low quality of satellite images, it is very hard to analyze any location whether an area has solar potential or not. Many people are also not aware of how much solar power their rooftop can generate. Our solution can provide an analysis of a rooftop remotely and it will motivate people to use more and more renewable energy rather totally dependent on the non-renewable sources of energy. We focused on the Adaptive Edge Detection and Contours to segment out rooftop boundaries and obstacles present inside them accurately. Our approach targets the placement of solar panels and provides an approximate analysis of the solar potential of a building or a house. Previous works done on solar panels placement was done mainly in foreign countries where they have 3D mapping too.

In the future, the authors aim to improve the approach by using Deep learning approaches to learn the substantial features. We are currently increasing our dataset for supervised classification of rooftop detection from Satellite images [18] using Mask R-CNN. [17] We are also considering rendering 3D depth reconstruction from a single image to improvise our analysis.[19] For improvement of edge detection and segmentation, we are currently exploring Gradient-based Edge-Aware filters [20] and Deep Active Contours[21] respectively.

References

1. Dong-Min Woo, Dong-Chul Park, Seung-Soo Han, and Quoc-Dat Nguyen, "Building Extraction Using Fast Graph Search".
2. Ildiko SUVEG, George VOSSELMAN, "3D RECONSTRUCTION OF BUILDING MODELS", TU Delft, Netherlands.
3. Antonis Katartzis and Hichem Sahli, "A Stochastic Framework for the Identification of Building Rooftops Using a Single Remote Sensing Image", IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 46, NO. 1, JANUARY 2008,

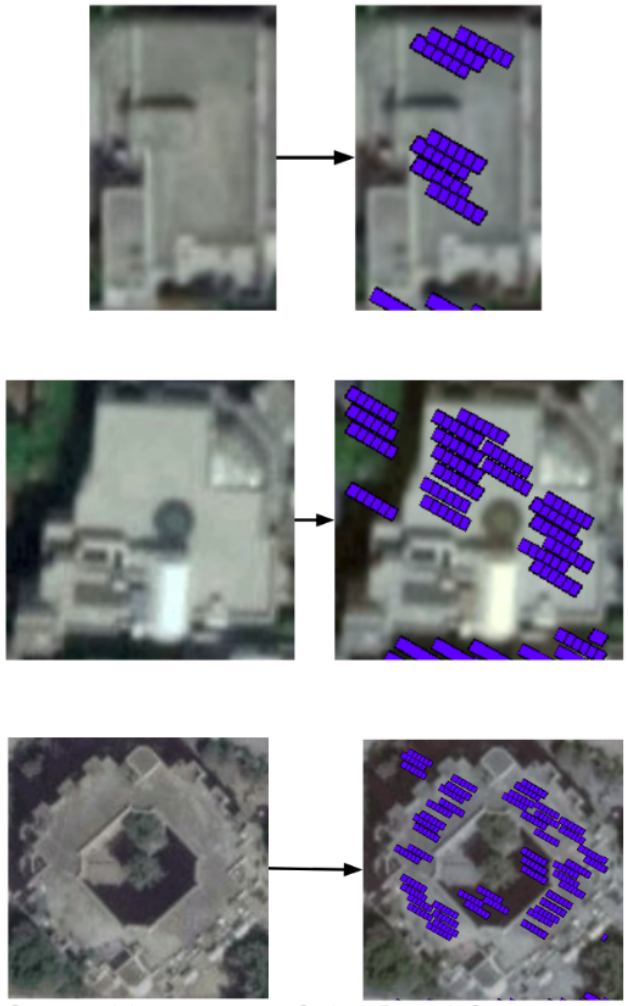


Fig. 8. Proposed Architecture output. The input image is on the left side and the blue solar panels area shows the optimal area available for solar panel placement with proper angle orientation.

4. Y. Huang, B. Yu, Z. Hu, J. Wu and B. Wu, "Locating suitable roofs for utilization of solar energy in downtown area using airborne LiDAR data and object-based method: A case study of the Lujiazui region, Shanghai," 2012 Second International Workshop on Earth Observation and Remote Sensing Applications, Shanghai, China, 2012, pp. 322-326. https://doi.org/10.1109/EORSA.2012.6261192
5. L. Zhang and X. Chen, "Topology-based automatic 3D modeling from multiple images," 2014 Sixth International Conference on Wireless Communications and Signal Processing (WCSP), Hefei, 2014, pp. 1-6. doi: 10.1109/WCSP.2014.6992055.
6. Gregoris Liasis, Stavros Stavrou, "Satellite images analysis for shadow detection and building height estimation", ISPRS Journal of Photogrammetry and Remote Sensing, Volume 119, September 2016, Pages 437-450. doi: 10.1016/j.isprsjprs.2016.07.006
7. K. Ni and Y. Wu, "Active contours driven by novel fitting term for image segmentation," in Electronics Letters, vol. 53, no. 13, pp. 854-856, 22 6 2017. doi: 10.1049/el.2017.1531.
8. Basu, S., Mukherjee, D.P., Acton, S.T.: Active contours and their utilization at image segmentation. In: Proceedings of Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence and Informatics Poprad., pp. 313317 (2007).
9. Airouche, M., Bentabet, L., & Zelmat, M. Image Segmentation Using Active Contour Model and Level Set Method Applied to Detect Oil Spills.Proceedings of the World Congress on Engineering 2009 Vol I WCE 2009, July 1 - 3, 2009, London, U.K.
10. Hossain, F., & Rahman, M.R. (2016). Dynamic Thresholding based Adaptive Canny Edge Detection. International Journal of Computer Applications (0975 8887) Volume 135 No.4, February 2016. doi: 10.5120/ijca2016908337
11. E. Pakizeh and M. Palhang, "Building detection from aerial images using Hough transform and intensity information," 2010 18th Iranian Conference on Electrical Engineering, Isfahan, 2010, pp. 532-537. doi: 10.1109/IRANIANCEE.2010.5507013.
12. M. Wang, S. Yuan and J. Pan, "Building detection in high resolution satellite urban image using segmentation, corner detection combined with adaptive windowed Hough Transform," 2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS, Melbourne, VIC, 2013, pp. 508-511. doi: 10.1109/IGARSS.2013.6721204.
13. Katiyar and Arun, "Comparative analysis of common edge detection techniques in context of object extraction," IEEE Transactions of Geoscience and Remote Sensing, vol. 50, pp. pp.68 2014.
14. G.T Srivanshan and Dr. C. Chandrasekar A Comparison of Various Edge Detection Techniques used in Image Processing International Journal of Computer Science & I.T (IJCSIT),Vol9,No 1,Sep 2012.
15. Palvi Rani and Poonam Tanwar (2013), A Nobel hybrid approach for edge detection, International academic paper of Computer Science & manufacturing assessment Vol.8, 827-938.
16. D. K. Chaudhary, R. Lal, N. Kashyap and T. Choudhury, "Hybrid edge detection technique for digital images," 2016 International Conference on Computing, Communication and Automation (ICCCA), Noida, 2016, pp. 1116-1121. doi: 10.1109/CCAA.2016.7813883.
17. He, Kaiming et al. Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV) (2017): 2980-2988.
18. Ghaffarian, S., & Ghaffarian, S. (2014). Automatic Building Detection Based on Supervised Classification Using High Resolution Google Earth Images.

19. Ashutosh Saxena, Sung H. Chung, and Andrew Y. Ng. 2008. 3-D Depth Reconstruction from a Single Still Image. Int. J. Comput. Vision 76, 1 (January 2008), 53-69. DOI=http://dx.doi.org/10.1007/s11263-007-0071-y.
20. Xu, L., Ren, J., Yan, Q., Liao, R. and Jia, J.. (2015). Deep Edge-Aware Filters. Proceedings of the 32nd International Conference on Machine Learning, in PMLR 37:1669-1678.
21. Rupprecht, C., Huaroc, E., Baust, M., Navab, N.: Deep active contours. arXiv preprint arXiv:1607.05074 (2016).
22. A. Zakharov, A. Tuzhilkin and A. Zhiznyakov, "Automatic building detection from satellite images using spectral graph theory," 2015 International Conference on Mechanical Engineering, Automation and Control Systems (MEACS), Tomsk, 2015, pp. 1-5. doi: 10.1109/MEACS.2015.7414937.

Step Up Gain Converter with fast MPPT control under moving partial shading for train rooftop PV-DC-µG

Shirish Raizada , Vishal Verma , *Member IEEE*
Electrical Engineering Department
Delhi Technological University,
Delhi, India
raizada.shirish@gmail.com , vishalverma@dce.ac.in

***Abstract*— A Dual Port Step-up Gain Converter (DPSC) fed rooftop PV train DC microgrid is proposed for operation under moving partial shading conditions. The proposed isolated converter provides high voltage gain using multiplier cells and galvanic isolation and thereby enabling interfacing of PV panel directly with 380V DC rooftop train DC-microgrid (DC-µG). The interleaving of the input stage along with low stress operation of devices at input stage makes the converter highly efficient. The DPSC delivers power to the DC microgrid by interfacing the two PV panels, one at each port without circulating current to form an element of the DC-µG. The fast dynamics of the converter along with high control bandwidth under extreme moving partial / full shading conditions due to moving objects are discussed in the paper. A fast MPPT algorithm is also proposed which tracks the maximum power point under moving partial shading conditions very fast and accurately. The proposed converter and algorithm exhibits fast dynamics with robust and reliable operation. The comparative analysis between the proposed control algorithm and conventional MPPT (P&O) is also presented in the paper.**

Keywords—Moving Partial Shading, Multiport, Voltage Multiplier, DC Microgrid. Rooftop PV Train

I. Introduction

There has been a steep rise in the greenhouse gases posing threat to the environment and health of human beings. The railway, worldwide contributes to 3.5% of the total greenhouse gases emission [1]. A major paradigm shift has been observed towards electrification of railway and electric traction. However, routing of power through pantograph and converter in traction system witness only around 60% of energy efficiency. One way to increase overall such efficiency of conversion is to offset the lighting and fan load in the moving train from alternative source of energy. PV panels presents an omnipresent low cost, low maintenance and compact size solution owing to its zero-greenhouse gas emission. The large roof of the railway cars of the train and railway station can allow large number of installation of panels and thereby offer ground for increased power generation, particularly in countries such as India, where abundant sunlight is available throughout the year.

The installation of rooftop photovoltaic panels on Lehrter Bahnhof railway station in Berlin and Takasaki railway station in Japan are some of the examples [2],[3]. The case studies dealt in analysis of the impact of the PV system in the metro line of Istanbul airport [4], whereas, the feasibility of generation is reported and installation of PV panels on the roof of the metro train in UAE is reported in [5]. In 2017, Indian railways launched their first solar train with rooftop PV panels on the coaches of the train [6]. The PV systems are shown to be very sensitive, and incur significant loss under partial shading conditions, a 20% shading on panels in railway is reported to reduce 60% of panel output power [7].

The PV panels are generally connected in series to form a string and the configuration is more popular due to the requirement of only nominal voltage gain from the interfacing converter [8]. But, partial shading on the panels of the string devastatingly affect the output voltage due to bypassing the partially shaded PV panels as the common current flowing in the string. Since there exist no possibility to find global maximum in a blink of time during moving partial shading on panels, the output is highly depreciated accompanied with oscillations in voltage at the output terminals. Recent researches are reported on one converter module per panel approach which evacuates more power from the panels as compared to the string configuration under moving / stationary partial shading conditions and offer simple techniques for finding out MPP to enable efficient operation of the system [9]. Furthermore, per panel converter module offers higher protection and allows redundancy of the power sources and converters. However, for fast variation in generated voltage a high voltage gain converter also becomes a requisite. The family of high voltage gain boost converters are reported in literature [10]. In the paper [11], the voltage multiplier cell based high gain converter is reported to offer wide input range of operation and modular architecture. The conventional approach of integrating different energy sources with train rooftop mounted microgrid requiring individual converters leads to higher component count and cost. Recently, the multiport converters are gaining popularity, as these can incorporate different sources at different port and offers compact and power dense solution [12]. The non-isolated multiport converters are reported in the literature [13],[14]. The interfacing of panel integrated converters for interface with DC microgrid at 380V or above requires converters with galvanic isolation for safety. The isolation issues of multiport converter between loads and sources are discussed in [15],[16].

The concept of rooftop PV panel on railway system for utility applications in the railway car is envisaged for development of systems and control. The compact and efficient distributed architecture of converter, interfacing several PV panels is proposed. The paper presents isolated dual-input-port converter incorporating multiplier cells with high frequency transformer based isolation. Multiple PV panels are integrated to the converter at different ports for operation in parallel with minimal circulating currents and feeding the power to the 380V DC bus. The salient feature of the proposed converter is its ability to maintain efficiency with high voltage gain and significant curtailment in the voltage stress across the devices. The converters allow the interleaving of the inductor current at the input to reduce the current ripple and in turn reducing the ripple in PV output power. The fluctuations in the power due to environmental conditions such as irradiation changes and the partial shading of the panels due to the fast-moving shadows of the trees and objects passing by the trains are deftly dealt. The control guarantees the extraction of maximum power from each PV panel under different conditions, even under moving partial shading. The control algorithm exhibits extremely fast dynamics with minimum computation which is a necessity for operation under moving partial shading. The effect on fast dynamics of the inductor current and voltage of multiplier capacitor under moving partial shading is also investigated in the paper. The effectiveness of the proposed converter system and control algorithm under moving partial shading on the portion of the panels, is demonstrated through simulations results in the MATLAB environment.

II. System Configuration

The new architecture of rooftop PV panels onboard railway cars is shown in Fig.1, where several PV panels are parallelly connected and then boosted to connect with 380V DC bus on the roof of the railway car. The parallel connection of two panels in group, coupled with isolated Dual Port Step-up Converter (DPSC) circuit feeding to the DC bus maintained by a master source at 380V. The proposed modular architecture allows maximum evacuation of power from the PV panels in comparison to the conventional architecture which comprises of string connected PV panels. The fast dynamics and distributed nature of architecture enables more accurate tracking of the PV power during moving partial shading for highly reliable operation due to its ability to operate at the reduced power capacity without fail as against centralized architecture.

This sectionalized architecture, which incorporates two PV panels in each section, and many such section aggregating on the rooftop of diesel-electric fed railway train car to constitute a rooftop train PV DC microgrid to supply the power to fan and light loads. The 500W section architecture is proposed in Fig.2. with two 250W PV panel support and such multiple sections feeding the DC bus. KD250GX-LFB PV panel is modelled with short circuit current of 9.09A and open circuit voltage of 36.9V with system parameters given in Table 1. The PV panels operating at the low voltage is interfaced with 500W DPSC so as to feed the power to rooftop train 380V DC bus. The two PV panels are individually connected at each input port of the DPSC to feed the high frequency transformer which output 380V peak voltage at its secondary. The interface converter at each input port allows

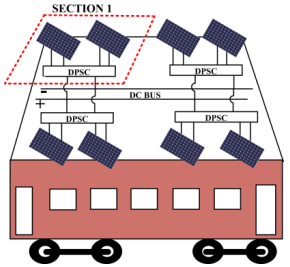


Fig.1. Proposed Sectionalized Architecture incorporating two PV panels per section on the rooftop train PV DC Microgrid.

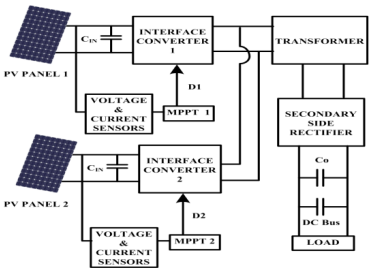


Fig.2. Section architecture representation of DPSC.

operation of individual PV panel near maximum power, providing immunity during moving partial shading on the panel. The faster dynamics of the interface converter (IC) of DPSC permits quicker tracking of the panels during moving partial shading.

III. DPSC Architecture

The converter topology of the proposed DPSC is shown in Fig.3, where two low voltage PV panels are integrated to the 380V rooftop train DC microgrid. For better understanding of the operation of DPSC, it is further considered fragmented into the input stage, multiplier stage, isolation-transfer stage and rectifier stage, wherein each interface converter (IC) for each PV panel consists of the input and multiplier stage. The input stage of each IC utilizes interleaved boost in conjunction with multiplier cells. The conduction losses are significantly reduced by the use of the low voltage high current OPTIMOS MOSFET having low R_{DSON}. The cross connected diodes along with the multiplier capacitor and blocking diodes helps in building the voltage and rejection of circulating currents respectively in the multiplier stage. The DPSC prohibits the accumulation of stress across the windings of the high frequency transformer, thereby allowing the linear operation of the transformer. The quasi square voltage tailored by synchronous rectifier appearing across the secondary winding

TABLE 1: System Parameters	
Maximum PV power	250W
Input Voltage (V_{MPP})	29.8V
Input Current (I_{MPP})	8.39A
Switching Frequency	50kHz
Turns Ratio	2
Multiplier Cell per interface converter (MC)	1
Inductor (L)	500µH
Multiplier Capacitance (C)	3.3µF

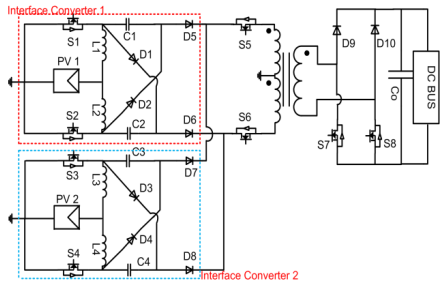


Fig.3. DPSC Architecture

978-1-5386-4192-7/18/$31.00 ©2018 IEEE

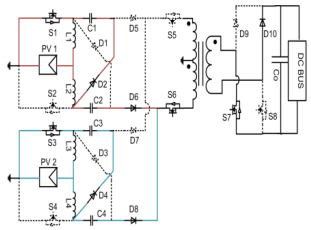


Fig.4.a Mode 1

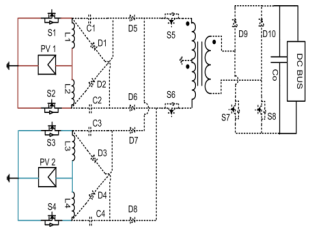


Fig.4.b Mode 2

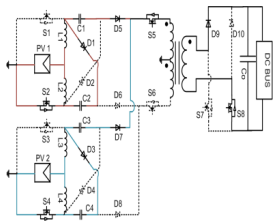


Fig.4.c Mode 3

of the high frequency transformer forms the rectifier stage to feed 380V DC bus.

IV. Modes of Operation

This section delineates the modes of operation of DPSC as shown in Fig.4.a-c. The proposed converter operates with operation as per input stage, multiplier stage, transfer switch cum isolation stage in tandem. The operation of individual IC is shown with both the PV panels operating identically under uniform insolation for the lucidity.

A. MODE 1:

Operation in this mode starts by turning ON of switches S_1 and S_3 and turning off of switches S_2 and S_4 of IC 1 and 2 respectively, thereby charging inductor L_1 and L_3, and discharging L_2 and L_4, along with the conduction of diodes D_2 and D_6 in IC 1 and D_4 and D_8 of IC 2. The aggregated currents from both the ICs flowing through switch S_6 is shown in Fig.4.a. in different colours. The sum of multiplier capacitor voltage in each IC appears across the primary winding of transformer. The switch S_7 and Diode D_{10} on the secondary side conducts to provide power to the 380V DC microgrid.

B. MODE 2:

Operation in this mode allows the charging of both input side inductor in each IC by turning ON the switches S_1 and S_2 of IC 1 and S_3 and S_4 of IC 2, allowing charging of both inductors L_1 and L_2 of IC 1 and L_3 and L_4 of IC 2. In this mode, none of the diodes of the IC 1 and IC 2 conducts as shown in Fig.4.b. Furthermore, both S_5 and S_6 are off, leading to no current in the primary winding and constructing a 3-level quasi square waveform on the transformer winding by turning ON of switches on the secondary side S_7 and S_8, thereby offering zero voltage across the transformer winding. The output converter capacitor continuous to provide the power to the DC microgrid.

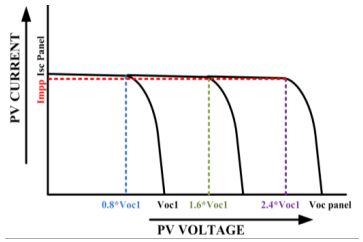


Fig.5.a I-V curve of the 3 sections of the PV panels

C. MODE 3:

Operation in this mode begins by turning OFF of switches S_1 and S_3 and turning ON of switches S_2 and S_4 of IC 1 and 2 respectively, hence charging the inductors L_2 and L_4, and discharging of L_1 and L_3, and conduction of D_1 and D_5 in IC 1 and D_3 and D_7 of IC 2. The summation of the different IC currents as shown in different colours flowing through switch S_5 is depicted in Fig.4.c. The aggregated voltage of multiplier capacitor in each IC appears across the primary winding of transformer. The switch S_8 and diode D_9 on the secondary side conducts to deliver power to the 380V DC microgrid.

V. Concept of Partial Shading

The PV panel suffers significant loss in the generation due to partial shading. The practical modelling of PV panels incorporates bypass diodes for each of the three sections (normal case) of the panel, which are connected in series. During the shading of the section of panel, diode activates the bypass to maintain the current flowing in other sections of the PV panel and thereby operate at overall lower panel voltage. Fig.5.a shows I-V curve of KD250GX-LFB, which is segmented into three different I-V curves for each of its individual sections. The open circuit voltage (V_{OC1}) of each section is 12.3V and its operation at maximum power is approximately 0.8 V_{OC1} for each section. Accordingly, the operation of second and third section of PV panel will be around 1.6 V_{OC1} and 2.4 V_{OC1} respectively. As the PV panel operates under uniform insolation of 1000W/m^2, it continues to track maximum power with the operation around 2.4 V_{OC1} and panel current at I_{MPP}. When the third section of the panel is considered getting shaded (black) as shown in Fig.5.b, diode activate bypass. For evacuation of maximum power from the unshaded portion of panel the operating voltage should be moved to 1.6 V_{OC1}, while maintaining the panel current at I_{MPP}. Moreover, even if the second section of the panel also gets full shaded, the operating voltage of the panel needs to drift towards 0.8 V_{OC1} to harvest remaining power from the panel.

VI. Control Algorithm for Moving Partial Shading

The rooftop PV panels onboard the train are partially shaded by the moving objects, thereby requiring control algorithm with fast dynamics. The conventional MPPT algorithms which work on highly down sampled rate take several steps to reach MPP, thus exhibits sluggish performance. Further the quest for global MPPT offer even more sluggish response reducing the power generated significantly from the PV panel under operation. In contrast, the proposed control algorithm (Fig.6) generates available power owing to its extremely fast convergence around

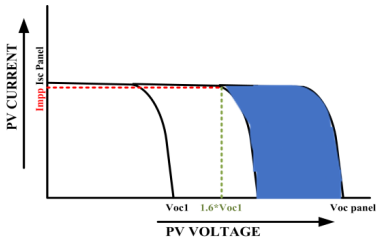


Fig.5.b I-V curve of the 3 sections of the PV panels with third section shaded

maximum power point under moving partial shading condition. The sensed panel voltage is given as input to the algorithm, which then compares it to the operating voltage points, i.e. 0.8V_{OC1} , 1.6V_{OC1} and 2.4V_{OC1} as discussed in the previous section. The algorithm then selects the duty cycle corresponding to the operating voltage and feeds it to the PWM module. The duty cycles for the particular operating voltages are calculated offline, which can be easily updated from time to time depending on the panel conditions. Moreover, when the panel voltage lies within the tolerance band around the operating voltage, the duty cycle remains constant. As the proposed algorithm operates between the three operating points, it ensures faster convergence near maximum power during moving partial shading.

VII. Analysis of DPSC

A comparative analysis dealing with voltage gain vs duty cycle for conventional boost converter and DPSC without the transformer gain (gain of 2) is shown in Fig.7. The conventional boost converter which witness nose dive drop in efficiency, the DPSC can sustain operation at about 95% efficiency even with higher duty ratio. The interleaving of the inductor current in the input stage reduces the input current ripple to an extent of 6%. The voltage across the multiplier capacitor of each module is maintained at 95V with 3.8% ripple in the voltage, thereby keeping the operation well within the standards. The multiplier capacitors accumulate a voltage of 190V appearing across the primary winding, making the secondary winding of the high frequency transformer to attain 380V voltage as depicted in Fig.8. The quasi square winding voltage at the output allows the reduction in the core size and core loss. The mathematical expression of the voltage gain for DPSC is shown in eq.1.where D signifies the duty ratio, V_O is output voltage, V_{in} is input voltage and N is the turns ratio of high frequency transformer.

$$M = \frac{Vo}{Vin} = \frac{2.N}{1-D} \qquad (1)$$

VIII. Performance Evalauation

The performance of DPSC fed from two PV panels under moving partial shading is presented in this section. Fig.9.a. shows the power drawn from terminals of panel 1 and panel 2 and the output power at the terminals of DPSC under the partial shading. For clarity both the panels are subjected to partially shaded conditions simultaneously at t=0.03s, when the power output from each panel getting reduced from 249W to 159W, in turn reducing the output power from 482 W to

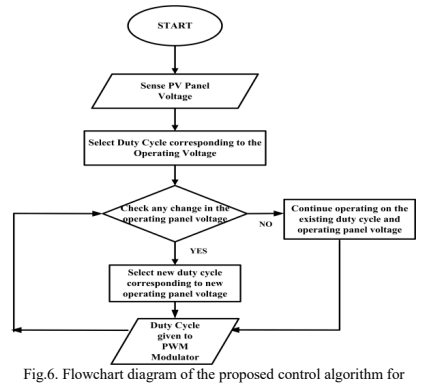


Fig.6. Flowchart diagram of the proposed control algorithm for moving partial shading.

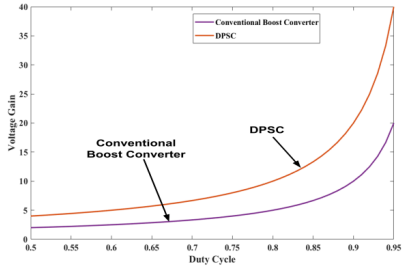


Fig.7. The comparative analysis of voltage gain vs duty ratio offsetting the transformer gain of 2.

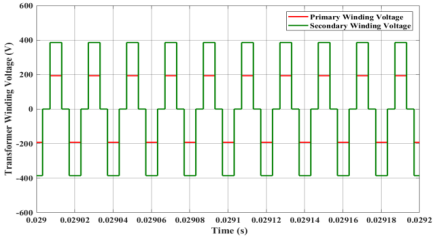


Fig.8. The primary and secondary winding voltage of high frequency transformer.

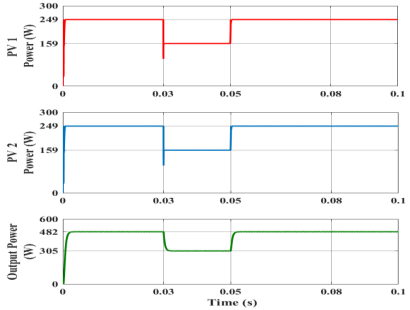


Fig.9.a. PV panel power and Output power under partial shading at 0.03s and removal of shade at 0.05s.

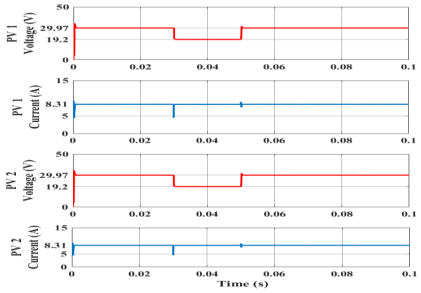


Fig.9.b. PV panel voltages and currents under partial shading.

305W. The removal of the shaded cover at t=0.05s with proposed MPPT algorithm enables the panels to quickly recover back to output 249W at their terminals, accumulating output power to 482W, exhibiting the fast convergence to MPP even under moving partial shading conditions. Fig.9.b shows the corresponding operating voltages and currents. Fig. 9.c depicts the performance of input stage coupled to panel 1 during transient operation where the removal of the shade at t=0.05s allows the panel voltage to recuperate back to 29.97V in 0.0003s, and the inductor currents and accumulated current at transfer stage also rapidly recovers back to steady state conditions. Furthermore, Fig.9.d depicts transient condition on the multiplier capacitor during the same transition; to operating from steady state voltage having small oscillations when the shade is removed and finally getting back within 3.5ms.

The performance of the proposed converter under moving partial shading is shown in Fig.10.a-d. Fig.10.a depicts power profile of the PV panels and the converter under moving partial shading. Panel 1 gets 1/3rd shaded during t=0.02s to t=0.035s, t=0.04s to t=0.05s and t=0.07s to t=0.08s and accordingly its power drops from 249W to 159W during the aforesaid intervals, while the panels 2 witness 1/3rd shading during t=0.06s to t=0.065s and t=0.07s to t=0.08s. It may be observed when one of the panel witness partial shading the output power drops to 395W, whereas furthermore, when both the panels gets shaded from t=0.07s to t=0.08s the output power drops to 305W from 482W. The dynamics of the electrical parameters, with, voltage and currents are evident from Fig.10.b. The fast dynamics of proposed converter and the control algorithm conforms to its merits particularly under moving shading conditions. The converter and the control algorithm complement each other and operates very fastly in tandem, and the same may be evident from Fig.10.b, where at t=0.02s when the panel 1 voltage drops to 19.2V the current dips a little but rapidly recovers back to 8.31A, and voltage and current of panel 2 remaining unaffected, signifying the ability of each module of the converter to operate independently and reliably without interference from neighbouring conditions. The dynamics of the inductor currents of each module are shown in Fig.10.c for same transient duration, where dip in current of panel 1 settles back to 4.155A in 0.0003s, while the inductor current of module 2 remains unaffected. Similarly, the multiplier capacitor voltage of module 2 remains unaffected whereas, for module 1 it takes 3ms to settle back to its position as shown in fig.10.d. The comparative superior dynamics of DPSC with proposed algorithm and that with conventional MPPT controller (P&O)

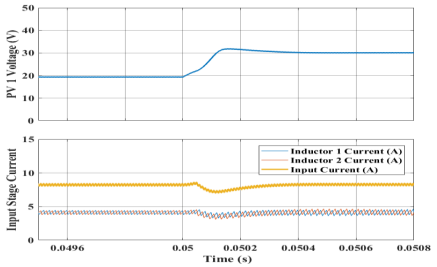


Fig.9.c. Transient of Inductor and Input Current and PV 1 Voltage when panel 1 partially shaded at 0.05s

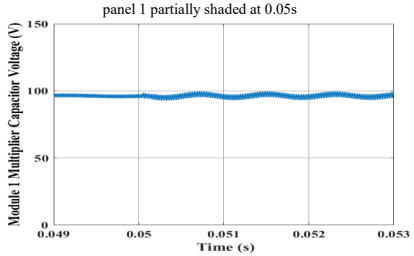


Fig.9.d.Transient of voltage of Module 1 Multiplier Capacitor at 0.05s.

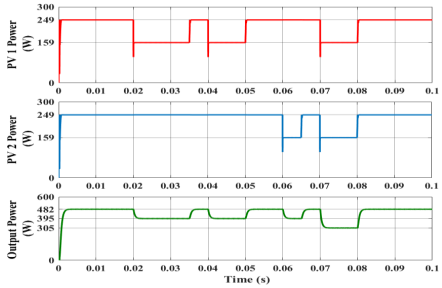


Fig.10.a. Power of PV panels and Output Power during moving partial shading

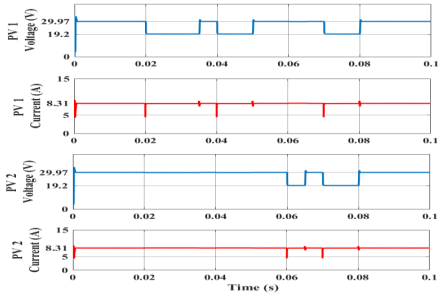


Fig.10.b. PV panel voltages and currents during moving partial shading

are depicted in Fig.11 under partial shading and its removal on panels. As the panel gets shaded at t=0.015s, the proposed controller tracks the maximum power of 159W of the partially shaded panel with minimal oscillations in 0.0005s, whilst the P&O gets stuck in the local optima, thereby settling to 139W

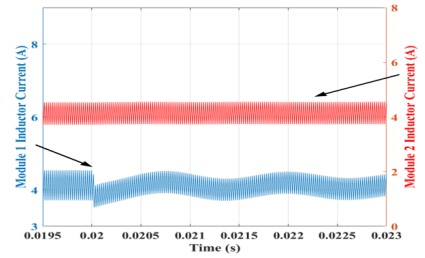


Fig.10.c. Transients of Inductor Current of both modules at 0.02s

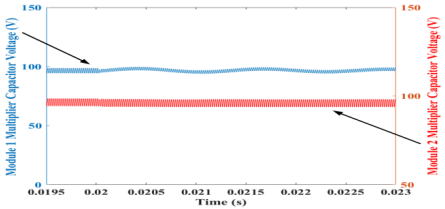


Fig.10.d. Transients of voltage of Multiplier Capacitor of both modules at 0.02s

in 0.0028s with large power oscillations, leading to fluctuating panel power and higher losses. Moreover, as the shading of the panel gets removed at 0.025s, the DPSC with proposed algorithm tracks the maximum power of 249W in 0.002s without significant oscillations in the power, demonstrating fast dynamic performance of the proposed algorithm with superior accuracy in tacking of panel power and minimal oscillations.

IX. Conclusion

A highly segmented and sectionalized train rooftop microgrid architecture is presented with new dual port high step-up gain converter (DPSC). The operation of the proposed DPSC is demonstrated with interleaving at the input stage for ripple reduction and interfacing of two panels without any circulating currents. The results demonstrate fast dynamics of multiplier cells as interface stage to offer quasi square voltage input to isolation transfer ensuring operation in linear range without saturation. The results demonstrate fast response of proposed converter with high gain and sustained good efficiency along with fast tracking of proposed algorithm

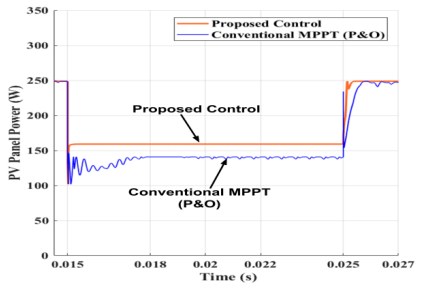


Fig.11. Comparison of performance of Proposed Controller and Conventional P&O MPPT Controller.

amidst moving partial shading conditions. The proposed control and architecture with proposed converter offer modularity and scalability for future expansion of PV-DC train microgrid under any condition of insolation and moving / stationary partial / full shading conditions by extracting all the available power from the panels.

References

[1] http://www.shop-etf.com/en/freepublications/railways-and-sustainable-development-a-globalperspective.html.

[2] "Photovoltaic Solar Energy Best Practice Stories", PV façade at railroad station Lehrter, European Commission, EC-contract no. SE-0220-96/D-UK.

[3] H. Hayashiya et al., "Necessity and possibility of smart grid technology application on railway power supply system," Proceedings of the 2011 14th European Conference on Power Electronics and Applications, Birmingham, 2011, pp. 1-10.

[4] B. Kilic and E. Dursun, "Integration of innovative photovoltaic technology to the railway trains: A case study for Istanbul airport-M1 light metro line," IEEE EUROCON 2017 -17th International Conference on Smart Technologies, Ohrid, 2017, pp. 336-340.

[5] V. K. Viswambharan, V. Kumar, K. Amjad and A. Ghani, "A case study — Solar powered metro in UAE," *2018 5th International Conference on Renewable Energy: Generation and Applications (ICREGA)*, Al Ain, 2018, pp. 46-51.

[6] http://www.financialexpress.com/economy/indiasfirst-solar-powered-train-all-you-want-to-know-about-indian-railwaysgreen-initiative/264505/.

[7] Y. Morita *et al.*, "Analysis of local smoothing effect on the PV on Tokyo station," *2012 International Conference on Renewable Energy Research and Applications (ICRERA)*, Nagasaki, 2012, pp. 1-6.

[8] E. H. Ismail, M. A. Al-Saffar, A. J. Sabzali and A. A. Fardoun, "A Family of Single-Switch PWM Converters With High Step-Up Conversion Ratio," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 55, no. 4, pp. 1159-1171, May 2008.

[9] W. Li, W. Li, X. Xiang, Y. Hu and X. He, "High Step-Up Interleaved Converter With Built-In Transformer Voltage Multiplier Cells for Sustainable Energy Applications," in IEEE Transactions on Power Electronics, vol. 29, no. 6, pp. 2829-2836, June 2014.

[10] M. Forouzesh, Y. P. Siwakoti, S. A. Gorji, F. Blaabjerg and B. Lehman, "Step-Up DC–DC Converters: A Comprehensive Review of Voltage-Boosting Techniques, Topologies, and Applications," in IEEE Transactions on Power Electronics, vol. 32, no. 12, pp. 9143-9178, Dec. 2017.

[11] Y. J. Acosta Alcazar, R. T. Bascope, D. S. de Oliveira, E. H. P. Andrade and W. G. Cardenas, "High voltage gain boost converter based on three-state switching cell and voltage multipliers," 2008 34th Annual Conference of IEEE Industrial Electronics, Orlando, FL, 2008, pp. 2346-2352.

[12] B. Karanayil, M. Ciobotaru and V. G. Agelidis, "Power Flow Management of Isolated Multiport Converter for More Electric Aircraft," in IEEE Transactions on Power Electronics, vol. 32, no. 7, pp. 5850-5861, July 2017..

[13] Hongfei Wu, Yan Xing, Yanbing Xia and Kai Sun, "A family of non-isolated three-port converters for stand-alone renewable power system," IECON 2011 - 37th Annual Conference of the IEEE Industrial Electronics Society, Melbourne, VIC, 2011, pp. 1030-1035.

[14] K. Tomas-Manez, A. Anthon, Z. Zhang, Z. Ouyang and T. Franke, "High efficiency non-isolated three port DC-DC converter for PV-battery systems," 2016 IEEE 8th International Power Electronics and Motion Control Conference (IPEMC-ECCE Asia), Hefei, 2016, pp. 1806-1812.

[15] W. Li, J. Xiao, Y. Zhao and X. He, "PWM Plus Phase Angle Shift (PPAS) Control Scheme for Combined Multiport DC/DC Converters," in IEEE Transactions on Power Electronics, vol. 27, no. 3, pp. 1479-1489, March 2012.

[16] W. Hu, H. Wu, Y. Xing and K. Sun, "A full-bridge three-port converter for renewable energy application," 2014 IEEE Applied Power Electronics Conference and Exposition - APEC 2014, Fort Worth, TX, 2014, pp. 57-

Study of Element Transfer for Silica Based Flux Constituents by Taguchi Analysis in Submerged Arc Welding

Aditya Kumar[1]
[1]Netaji Subhash Institute of Technology, Sector-3, Dwarka, New Delhi
(E-mail: aditya_rathihere@yahoo.com)

Abstract—A silica-based flux system was designed consisting of SiO_2-Al_2O_3-CaO. Taguchi L 9 orthogonal array was used to design the flux system for submerged arc welding. Three constituents of flux were varied, with their three levels such as, NiO, MnO and MgO. Open circuit voltage was kept as 32V, nine experiments were performed for bead on plate arrangement. Element transfer study was analyzed for silica-based flux constituents. Weld carbon content increases because of low oxygen content. This happens due to presence of Mno in the flux. Silicon content decreases in weld metal because of CaF_2 present in the flux. Copper content also decreases in the weld metal. Sulphur content increases, whereas phosphorus content decreases, because of transfer into the slag during slag metal reaction.

Keywords— Element transfer**,** Δ Quantity, SAW Flux Design **,**Basicity Index,

I. Introduction

Submerged arc welding is the process that consume electrode along with the agglomerated flux. The composition of the flux plays an important role to make a welded portion, stronger in terms of mechanical strength. During the welding process flux converts into weld metal and the slag. The slag covers the weld zone and it serves two purposes one is to protect the weld zone from atmospheric oxidation and another it also acts as heat barrier, to provide the better penetration. The main constituent of the flux is SiO_2. The main aim of silicon die oxide is to enhance the viscosity of the flux [13]. Fused flux yields less oxygen in weld deposit then do agglomerated fluxes with the same constituent's. Al addition to the flux lower the oxygen content of weld deposit. 1.2% of Al improves the weld notch toughness [3]. Calcium silicate and magnesium silicate fluxes shows a strong interaction with Cr and Si transfer and no interaction with Mn. Manganese silicate flux produces lower residual sulphur while calcium silicate flux is more effective for removal of phosphorus [4]. In a CaO-Al_2O_3 based flux system, source of oxygen and nitrogen contamination has been made to study and the result shows the major source of nitrogen contamination is by reaction with atmosphere at the electrode tip. However, the oxygen contamination is associated not only with reaction with atmosphere but also with flux decomposition and flux entrapment [6]. In a CaO-MgO-CaF_2-Al_2O_3 flux system, the result shows that individual flux ingredients and their binary mixture have a predominant effect on the weld metal transfer of oxygen, Manganese, Silicon and Carbon content [7]. The composition of flux and combination of metal-core wire has a significant effect on tensile strength and impact strength of the weld metal in terms of element transfer between the slag and weld metal. Both Carbon and Manganese shows negative Δ quantity in most combination. More basic flux yield less negative Δ C and Δ Mn [8]. A cost effective agglomerated flux was developed using flux dust the result shows that chemical composition and mechanical properties of all weld metal prepared from developed flux and parent flux to be in the same range [9]. An Optimization method of stastical design of mixture experiments with mathematical programming optimization for different flux constituents in submerged arc welding has been proposed [10]. Review on the flux composition and its behavior on bead geometry is presented in this study for submerged arc welding. [11]. Slag is recycled and replenished with suitably alloying elements and deoxidizers by agglomeration method. The result shows that the flux developed by recycled slag is with the accepted range of AWS [12]. The oxygen level in submerged arc welding is controlled by SiO_2 decomposition in most acidic flux, whereas for basic flux oxygen level is controlled by oxygen potential of slag determined by FeO contents of the slag [13]. All the above literature review suggests to design a flux from these constituent's combination to have good strength and machining properties.

II. Design of Flux by using Taguchi Method

Taguchi methodology is a simple and systematic approach of experimental design that can reduce number of experiments to an optimal level. In present work taking SiO_2 as the base of the flux three alloying elements NiO, MnO and MgO were varied at three different level. Proper selection of orthogonal array is based upon degree of freedom. Degree of freedom for each factor is one less than level of a factor. So, in present design, total degree of freedom for all three factors becomes 6. Since 2 degrees of freedom is assigned for each three factor, and 2 degrees of freedom is assigned to error term, so total degree of freedom becomes 8. Number of experiments can never be less than the total degree of freedom. Keeping all this in view, L9 orthogonal array was selected which has nine numbers of runs. Nine experiments were run for SiO_2 based flux by varying alloying elements in different combination as per the design.

TABLE I. LEVEL OF FACTORS

Factors	Code	Level 1	Level 2	Level 3
NiO (Wt in gms)	A	60	80	100
MnO (Wt in gms)	B	50	70	90
MgO (Wt in gms)	C	85	95	105

TABLE II. COMPOSITION OF SiO_2 BASED FLUX

Constituents	Al_2O_3	TiO_2	CaO	CaF_2
(Wt %)	18.45	28.0	27.15	7.5

III. Preparation of Flux

Based on the above analysis and selection of orthogonal array, 9 fluxes having SiO_2 as base were prepared. The Main three constituents that vary in the flux were NiO, MnO and MgO, their level is shown in Table 1. Other constituents of flux that were kept constant are shown in Table 2. Total 2kg of flux was prepared for each combination by the agglomeration method. All the constituents were dry mixed first in a big container with ten steel balls of 10mm diameter for half an hour. Potassium silicate was heated up to 50^0C, sprayed into the mixture (1000ml per 500ml of water). Same mixture was then thoroughly mixed again for half an hour. This process was repeated again and again until the binder potassium silicate and other constituents have formed the granules. After that the prepared mixture was dried at room temperature for 48 hours and then it is baked in muffle furnace at approximately 650^oC for 4 hours. After normal cooling, the prepared mixture was passed through the sieve of required size to remove the dust from the flux and kept in air tight containers. The flux was baked again before use to eliminate any possible moisture.

IV. Experimental Procedure

A constant DC voltage submerged arc welding machine was used for welding the mild steel plate of 300 x 150 x 20 mm using 3.15mm diameter wire of grade EL8 DIN8557: SI by beveling the plate edges (forming an angle of 45^0 in between). The machine had the provision for automatic wire feed rate and trolley speed. A backing plate of 10mm with same material grade was attached to the main plates. The welding parameters, as shown in the Table. 3, were kept constant for each experiment**.**

TABLE III. WELDING CONDITION PARAMETER

Parameter	Unit	Value
Voltage	V	32
Trolley Speed	cm/min	20
Electrode stick out		2.5mm.
Flux		SiO_2 based

Three to four-layer weld passes were made for filling the groove in each experiment for 9 fluxes of SiO_2 base. After sectioning the plate, at center nine samples were taken to study the element transfer. Small holes are drilled in the plates to take the chips. These chips are collected in separate boxes to test the chemical composition of the chips the results of test are shown in the Table.IV

V. Result and Discussion

A. Calculation of Δ Quantity

The Δ Quantity of the flux system is a measure to find the quantity of elements transferred to the weld metal or the slag. The Δ Quantity can be calculated by the following equation. The same was used by Dallam et.al (1985) in slag metal reaction.

Δ Quantity = Analyzed composition – Expected composition

The welded plate composition was analyzed by the chemical composition of the weld metal and it is shown in Table.IV
In the equation {1} shown below base plate composition means the quantity of elements present in the base plate and wire composition means the quantity of element present in the wire composition. 10 % of the base plate composition and 90 % of the wire composition was considered to be transferred into the weld metal. The Δ Quantity can be negative and positive. If it is positive that indicates that the element has been transferred from flux to the weld metal, if negative it means that the element has moved to the slag from the flux. Using the above equation, the Δ Quantity was calculated and tabulated in the Table.V relative to each flux.

TABLE. IV EFFECT OF ELEMENT TRANSFER OF SiO_2 BASED FLUX AT 32V

Flux No	Elements % age in welds at 32V with SiO_2 based fluxes							B.I OF Flux
	C	Si	Mn	S	P	Cu	Fe	
111	0.1	0.12	0.64	0.01	0.018	0.42	98.692	1.133
122	0.2	0.12	0.5	0.006	0.018	0.12	99.036	1.161
133	0.2	0.15	0.35	0.014	0.017	0.011	99.258	1.189
212	0.22	0.12	0.3	0.006	0.018	0.11	99.226	1.147
223	0.18	0.1	0.48	0.006	0.016	0.14	99.078	1.175
231	0.22	0.13	0.3	0.004	0.017	0.1	99.229	1.161
313	0.21	0.13	0.4	0.005	0.017	0.13	99.108	1.161
321	0.2	0.14	0.49	0.005	0.018	0.1	99.047	1.147
332	0.11	0.11	0.62	0.009	0.019	0.26	98.322	1.175

$$\text{Expected composition} = \frac{\text{Dilution X Base plate composition} + (100 - \text{dilution}) \times \text{wire composition}}{100} \quad \text{-\{1\}}$$

Basicity Index = Basic Oxide / Acidic Oxide ---{2}

TABLE. V Δ QUANTITY SiO_2 BASED FLUX

Flux No	32v Sio2					
	Δc	ΔSi	ΔMn	ΔS	ΔP	Δ Cu
111	-0.0099	-0.0098	0.244	-0.013	-0.01	0.234
122	0.091	-0.0098	0.104	-0.017	-0.01	-0.065
133	0.091	0.0202	-0.046	-0.009	-0.011	-0.174
212	0.1101	-0.0098	-0.096	-0.017	-0.01	-0.075
223	0.0701	-0.0298	0.084	-0.017	-0.012	-0.045
231	0.1101	0.0002	-0.096	-0.019	-0.011	-0.085
313	0.1001	0.0002	0.004	-0.018	-0.011	-0.055
321	0.091	0.0102	0.094	-0.018	-0.01	-0.085
332	0.0001	-0.0198	0.224	-0.014	-0.009	0.074

The main objective of the study was to study the transfer of elements from flux into weld metal and slag. The voltage condition was kept constant at 32 V. The flux constituents were varied in different proportions to find the effect of each constituents based on the basicity index of the flux as shown by equation {2}. Basicity index of the flux decide the nature of the flux whether the flux is acidic, basic or neutral.

The variation of Mn "Fig.1" shows that at 32 volt, the Mn content varies from 0.3% to 0.64%, The possible reason may be the transfer of Mn to the slag. The Mn reacts with SiO_2 and forms Manganese silicate, which may be transferred to the slag because it is more soluble in the slag. The second reason may be the evaporation of Mn as it is vaporized easily.

The carbon content in the weld increases hardness and tensile strength but it is not supposed useful for impact strength. "Fig.2" shows that the weld carbon content increases from 0.1% to 0.22% with increasing B.I at 32 volts. This may be because of low oxygen in the welds as given by (Bang et.al 2009) because ΔC is the function of oxygen content.

The weld Silicon content reduces in the welds with the CaF_2 present in the flux as shown in "Fig.3" This CaF_2 may form SiF_4 gas as said by (Tulani.et.al 1969). Si content increase with the increase of Al_2O_3 because Al is a strong deoxidizer (Bang et.al 2009). In our case the Al_2O_3 ratio is fixed in all the fluxes. According to (Davis and Bailey et.al.1980), an increase in Al, Ti, and Zr could replace Si in the SiO_2 network, leaving more Si free to transfer from the slag to the weld metal because Al, Ti, and Zr oxides are also network-formers like SiO_2.

SiO_2 fluxes at 32 volts, the weld sulphur content shows an increasing trend as shown in "Fig.4". This is because individual flux ingredients CaO, MgO, CaF_2, and Al_2O_3 are responsible for the gain of sulphur in the weld metal. This may be because of reduction in oxygen content with increasing B.I of the flux. Sulphur is removed by reduction reaction (U. Mitra et.al 1984, 1991), therefore these three ingredients will help the gain of sulphur in the weld metal. However, MgO was not reported to increase weld metal oxygen content. In spite of that, the gaining of sulphur by MgO indicates that it is not capable of desulphurization (Oiks, G et.al 1977). All these phenomena probably reduce the sulphur absorbing capacity of MgO in the flux.

Phosphorus reacts with the availabe oxygen released by acidic oxide and form phosphorus oxide or phosphorus penta oxide. All the flux of SiO_2 bases contais the CaO and MnO these two reduces the phosphorus in the weld metal (U.Mitra 1984). All the ΔP quantities in SiO2 based flux are negative so it is presumed that most of the phosphorus is transferred to the slag.

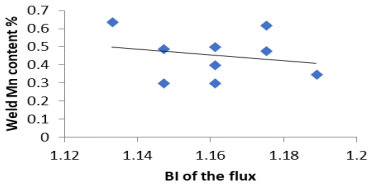


Fig.1 Weld Mn Transfer

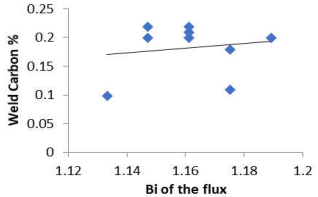


Fig. 2 Weld Carbon Transfer

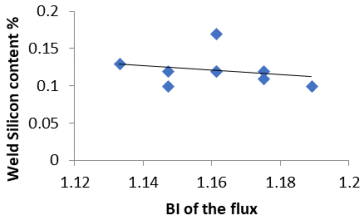


Fig. 3 Weld Silicon Transfer

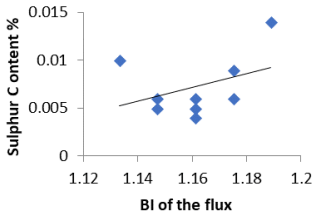


Fig. 4 Weld Sulphur Transfer

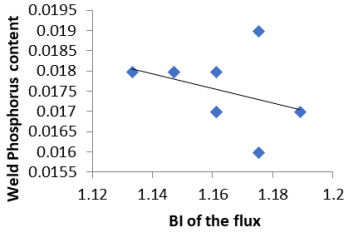


Fig. 5 Weld Phosphorus Transfer

REFERENCES

1. Tulani, S.S, Boniszewski, T, Eaton, N.F., "Notch Toughness of Commercial Submerged arc weld Metal.Welding and Metal Fabrication" (1969), Vol.37(8), pp- 327-339.
2. Oiks G, "Converter and open-hearth steel manufacture Fundamentals of theory and Technology" (1977), Mir publisher Moscow: 97-200,324-334.
3. T.H.North, H.B.Bell, A.Nowicki and I.Craig. The welding journal. 1978. pp - 63-s- 75s,
4. Davis M.L.E. and N. Bailey, "Proc. Conf. on Weld Pool Chemistry and Metallurgy". The Welding Institute, London, UK (1980), 28
5. U.Mitra and T.W.Eagar Metallurgical transactions A, Vol.15A, ,1984. pp:-217-226
6. P.Kanjilal,T.K.PalandS.K.Majumdar Welding Journal . Vol.- 86,pp:135s-145s.2007
7. P.Kanjilal. T.K.Pal and S.K.Majumdar, " Predication of element transfer in submerged arc welding" Welding Journal . Vol.-86. 2007, , pp: 135s-145s
8. Kook-soo Bang, Chan Park, Hong-chul Jung and Jong-bong Lee , "Effect of flux composition on the element transfer and mechanical properties of weld metal in submerged arc welding" Met.Mater.Int. Vol-15 .pp: 471-477 , 2009.
9. Vinod kumar,Dr. Narendra Mohan and Dr. J.S.Khamba . "Development of cost effective agglomerated fluxes from waste flux dust for submerged arc welding", Proceedings of the world congress on engineering .Vol-1, 2009.
10. Ademola David Adeyeye and Festus Adekunle Oyawale,"Weld-Metal property optimization from flux ingredients through mixture experiments and mathematical programming approach", Materials research , Vol. 12, 2009, pp: 339-343,
11. Brijpal singh, Z.A.Khan and A.. N.Siddique, "Review on effect of flux composition on its behavior and bead geometry in submerged arc welding", Journal of mechanical engineering research , Vol.-5, 2013, pp: 123-127,,
12. Kulwant singh, V.sahni and S.Panday ," Slag recycling in submerged arc welding and its influence on chemistry of weld metal", Asian Journal of chemistry, Vol. -21, 2009, pp: -SO047-051,
13. T.W.Eagar, " Source of weld metal oxygen contamination during submerged arc welding", Welding journal research supplement . 1978.



Aditya Kumar was born in Mawana, Distt. Meerut, U.P in 1967 India. He obtained his B.E degree from K.L.E.S. Engineering College, Belgaum, Karnataka in the year 1993 with Mechanical Engineering. He obtained his Master's degree from Delhi College of engineering in the year 1999. He joined as Senior Scientific Assistant at Netaji Subhas Institute of Technology in Feb 1994 and got the promotion as Lectures in year 2002 and subsequently as Assistant Professor in year 2007 and Associate Professor from 2015 onwards, he has a long teaching experience at NSIT about 24 years and published 8 papers in International Journal, 6 papers in International and National Conference. He is a life member of ISTE and IIW.

emeraldinsight

Benchmarking: An International Journal

Technology forecasting (TF) and technology assessment (TA) methodologies: a conceptual review
Abid Haleem, Bisma Mannan, Sunil Luthra, Sanjay Kumar, Sonal Khurana,

Article information:

To cite this document:
Abid Haleem, Bisma Mannan, Sunil Luthra, Sanjay Kumar, Sonal Khurana, (2018) "Technology forecasting (TF) and technology assessment (TA) methodologies: a conceptual review", Benchmarking: An International Journal, https://doi.org/10.1108/BIJ-04-2018-0090
Permanent link to this document:
https://doi.org/10.1108/BIJ-04-2018-0090

Downloaded on: 05 January 2019, At: 22:00 (PT)
References: this document contains references to 101 other documents.
To copy this document: permissions@emeraldinsight.com
The fulltext of this document has been downloaded 4 times since 2018*
Access to this document was granted through an Emerald subscription provided by emerald-srm:544911 []

For Authors

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com

Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

*Related content and download information correct at time of download.

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

The current issue and full text archive of this journal is available on Emerald Insight at:
www.emeraldinsight.com/1463-5771.htm

Technology forecasting (TF) and technology assessment (TA) methodologies: a conceptual review

Abid Haleem and Bisma Mannan
Department of Mechanical Engineering, Faculty of Engineering and Technology, Jamia Millia Islamia, New Delhi, India
Sunil Luthra
Department of Mechanical Engineering, State Institute of Engineering and Technology, Nilokheri, India
Sanjay Kumar
Department of Mechanical Engineering, Dr Akhilesh Das Gupta Institute of Technology and Management, New Delhi, India, and
Sonal Khurana
Department of Mechanical Engineering, Faculty of Engineering and Technology, Jamia Millia Islamia, New Delhi, India

Received 5 April 2018
Revised 10 May 2018
Accepted 13 May 2018

Abstract

Purpose – Technology forecasting (TF) and assessment (TA), all in all, apply to any intentional and deliberate endeavours to forecast and view the potential heading, rate, attributes and impacts of technological change, especially for development, advancement, selection and utilisation of resources, which ultimately helps in the benchmarking. A vast variety of methods are available for TF and TA. Till now, practically, no exertion has been made to choose proper, satisfactory innovation methods or technology. The paper aims to discuss this issue.
Design/methodology/approach – In this paper, there is an endeavour to summarise the vast field of TF and TA, through its evolution, functions, applications and techniques. This paper provides the in-depth review of the utilisation of TF and TA methodologies and its improvement, which helps the users in selecting the appropriate method of TF and TA for a specific situation.
Findings – This study concludes that the quest for a single strategy for doing forecast and assessment is a misconception. This neglects to perceive that forecast and assessment oblige a suitable blend of strategies and methods drawn from a variety of fields. Researchers and practitioners must be innovative, imperative and specialised in choosing TF and TA methodologies, and cannot be programmed.
Practical implications – The technology seems to be the most significant driver of the present day global developments. Some technologies have far-reaching implications, and the authors need to understand these issues regarding its' forecasting and its assessment.
Originality/value – The decision of proper worthy procedure amid a circumstance may have an impact on the exactness and reliability of the forecast and assessment. Significant observations regarding learning, action/s, actor/s and expected outcomes are discussed.

Keywords Benchmarking, Literature review, Technological forecasting, Technology assessment, TF and TA methodologies, Techniques of technology forecasting

Paper type Literature review

1. Introduction

Business organisations have been implementing modern technologies to encounter present challenges towards new/better products, processes/activities, services and practices for delivering higher efficiency and effectiveness (Kumar *et al.*, 2015). Analysis of the implications of new or emerging technologies has become a necessity for the economies,

Benchmarking: An International Journal
© Emerald Publishing Limited
1463-5771
DOI 10.1108/BIJ-04-2018-0090

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

organisations and societies. This analysis does not only help in cost-effective solutions, but is also vital for identifying the best suitable alternative ranging from individual level to the global level. These analyses are the methodologies of technological forecasting (TF) and technology assessment (TA) that help in making a well-informed decision and setting priorities for research and development (R&D). TF and TA also help in exploiting the available knowledge and understanding, as well as managing the risk of emerging technologies and innovations. As the business environment is extremely competitive in public as well as private sector, TF and TA are used to satisfy the rising needs of the innovation (Huang *et al.*, 2014). Generally, a country's economy depends upon the innovations to keep up and enhance living standards. In today's global world, countries are interconnected for the advancement of the economy (Bijker, 1992). Along these lines, TF and TA are more imperative and hard to target for investigations to help decision maker (Ely *et al.*, 2014). Technological forecasting and evaluation will generally rely on the perspectives of the overall population and their constituent agents about financial competition, technological advancement and the technological improvement (Jun, 2011; Compagna and Kohlbacher, 2015).

The best analogy for TF and TA is climate forecasting. This is a brilliant TF, which permits, arranges, decides and encourages expanding gain and minimising losses for the future conditions. In the case of climate forecasting, people verifiably figure the climate by their option of whether to wear a waterproof coat, convey an umbrella or put the sunscreen, etc. (Mishra *et al.*, 2002).

Any individual, or association, or country that may be influenced by the change in technology takes part in forecasting and assessment of technology in every sphere that dispenses assets to express purposes (Un and Price, 2007).

Benchmarking is a technique that compares the performance outcome of an organisation with the available technologies and the best practices of other organisations (Tasopoulou and Tsiotras, 2017). The application of the benchmarking is to focus on the identifying and understanding the detailed organisation procedures and after that comparing the organisation performance with that of other organisations (Guimaraes and Langley, 1994; Moriarty and Smallman, 2009). Managerial teams conduct the process of benchmarking using the results of TF and TA (Adebanjo and Mann, 2000; Daim and Dash, 2011). The first step of benchmarking is to identify and assess the available technologies and the outcomes of the organisation. In the next step, benchmarking is done using and comparing the available TF results or comparing the organisation performance with the competitive organisations. The shortcomings in the process or outcomes of the organisation are assessed using TA. In this way, a performance gap is established, and the elements, which have led to superior performance, are understood (Sharif, 2002). The final step is to formulate an improvement plan and implement the actions necessary to close the performance gap.

1.1 Objectives of the research

More than 80 per cent of the organisations are of the view that TF and TA are unstructured and unsystematic processes, and there is a need for improvement (Reger, 2001; Firat *et al.*, 2008; Kerr and Tindale, 2011). In this paper, an attempt has been made to identify respective tools for TF and TA. Both are the multidimensional processes and highly subjective. Many of the tools used for TF and TA may come with some customisation. This review work helps in exploring the objectives, as follows:

- to explore the historical evolution of TF and TA;
- to identify respective tools/methodologies for TF and group them under families;
- to analyse relationships among TF methods to explore limitation of these identified families; and
- to identify respective tools/methodologies for TA.

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

1.2 Methodology

The organisations use the number of different tools and methods with varying intensity for TF and TA. They do not have a clear idea about the distinctive method of TF and TA and where to use the TF method and where to use the TA at what intensity. From a methodological point of view, the literature review is comprehensive, and its content analysis was undertaken. We go through papers of over a wider horizon from 1960 to 2015. After going through all the available papers on Web of Science, Scopus and universities press sites, TF and TA methodologies and applications are summarised.

This study addresses numerous overlapped forms of forecasting technology and assessment methods and their effects. First, this paper offers the list of different methods of TF and then categorises them into eight major families. After that, the in-depth review of TA and its methodologies are presented. Authors' endeavour is to provide suggestions on the applicability of different methodologies as the case requirements. This paper gave a brief rundown of the groups of TF and techniques of TA and bolstered that the forecasting and assessment incredibly rely on upon fitting the choice and utilisation of proper technique. This paper also suggests that TF and TA are multidimensional activities.

We have organised the paper into seven sections. Section 2 deals with the TF and TA evolution over time. Section 3 presents various methods of TF, whereas Section 4 deals with the analysis of the relationship among identified TF methods and systematic analysis of limitations of TF followed by the need of TA and description about tools for TA in Section 5. Sections 6 and 7 present the discussion of findings clearly showing managerial implications and unique contributions of the paper and concluding remarks, respectively.

2. The evolution of TF and TA

The evolution of both (TF and TA) has been divided into two parts, i.e., first, before and, second, after the Second World War. Before the Second World War, the essential authority record of an exploratory point of view towards the long running of innovation and advancement of science which happened in 1935. National Resources Committee (1937) report suggested fundamentally about the outcomes of technological changes (Schnaars, 1989).

After the Second World War, the USA tried to investigate post-war effects mostly with the diagrams of statistical approach (Bush, 1990) and the USA focussed absolutely on the pace of innovative correction and technological change after the Sputnik stun. This accentuation likewise increases the needs of the US defence system where they can be focussed on progressive eras of weapons frameworks. As an outcome, the R&D administration system has been formed in both industry and government, which helps in setting up labs, funding research and building up research group (Roussel *et al.*, 1991). War yielded the need to address sensational improvements in innovation like guided rockets, atomic weapons and figuring. Framework investigation turned into a fundamental device in arranging such propelled frameworks. By 1949, basically, underneath the US Government, the improvement of TF as an exploratory method for investigating the eventual fate of innovation was underway. TF helped strategists of the military to manage the quality and lead times important to create an edge. Quantitative exploratory studies are based on taking data from the past and predict the future. It includes the development of models and pattern extrapolation. Regularising estimate, beginning with perceived future needs, is assumed a part too. The blend

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

likewise encouraged more subjective methodologies, for example, significance trees (Esch, 1972), mission stream examination, situation composing and Delphi (Gordon and Helmer-Hirschberg, 1964).

TF and TA were constructed to support industry, government and the educated community (Jantsch, 1967). In the 1960s, books and journals on TF and TA reached out a long-way past the US defence system. In the 1970s, disillusionment with frameworks investigation starts to spread with the disappointment of its capacity to manage ill-organised frameworks (e.g. Vietnam). In 1973, oil-stun uncovered a percentage of the impacts of geopolitical dangers on the forecast of innovation fates.

Hence, the support for TF and TA in policymaking has been started in the 1980s as it was understood that the instabilities of technology advancement challenged obvious "system analysis" arrangements.

There was additionally developing acknowledgement started a while ago that acknowledged scientific paradigm procedure, for example, approval and replication could not be connected to affirm the forecasting tool beyond the close term (Zhu and Porter, 2002). These are worries that will be enlightened in the discourses of numerous points of view and started the segment on the methodologies and tools for TF and TA. Technology-intensive firms with started to shift towards decentralised R&D administration and the second era of R&D administration systems. TF and TA reduce the practise in a set of tools; the use of forecast and assessment somewhere around 1975 and the mid-1990s was generally very less, for the most parts were ineffectively characterised and executed without much thoughtfulness regarding formal presumptions, time horizons or impediments (Coates *et al.*, 2001).

From the evolution of TF and TA, we may conclude with the following observations.

Observation 1: TF and TA may be recognised as an important area that needs to be explored in terms of tools/methodologies of TF and TA towards achieving business viability, securing profitability, and gaining competitive advantage for firm/supply chain and helping the human race by providing innovative and sustainable products/services/solutions.

Observation 2: however, some literature on the evolution of TF and TA is available, but there is a strong need to identify and analyse TF and TA tools/methodologies relevant to new product development (NPD).

3. TF methods

Settling on vital choices for TF method is a standout amongst the most troublesome difficulties for the R&D staff. As Ralph Lenz, US aviation-based armed forces innovation forecasting pioneer once said, "TF may be characterised as the prediction of the development, qualities, measurements, or execution of a machine serving some valuable need" (Slocum and Lundberg, 2001; Martino, 2003). Mishra *et al.* (2002) said that "Emphasis on TF method" and "Development of methodology to select an appropriate technique for TF is vital requirement of sustainable development".

3.1 TF families and associated methods

There are numerous methods of TF, which can be fit in under eight families (expert opinions, trend analysis, monitoring and intelligence methods, statistical methods, modelling and simulation, scenarios, value/economic/decision methods, descriptive and matrices method). Table I explains these identified TF families and the associated methods.

The TF methods have been identified from the relevant literature and fitted into eight families by utilising experts' advice. From Table I, we may conclude with the observation and sub-observations as provided below.

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

Table I.
An overview of TF families and related TF methods

TF family and TF methods	Brief description	Conditions for using TF methods
1. Expert opinions Focus groups (workshops, panels) Delphi Participatory techniques Interviews	Expert opinion incorporates forecasting or comprehension innovative improvement through serious counsel with specialists. In this group, the most prevalent strategy is the Delphi method. Delphi technique consolidates expert conclusions. It concerns with the probability of understanding the technology, which is proposed in the work of Haque *et al.* (2013). In this technique, there is a grouping of personals suggestion is trailed by supposition criticism got from dissecting the introductory reaction data. The feedback which is incorporated in the thought process or support which is behind every personal will help the experts in forecasting. It permits other experts to modify their forecast by considering the new data (Levary and Han, 1995). Delphi is being the most broadly utilised procedure and has been subjected to examination by numerous creators. Woudenberg (1991) has talked about the precision and dependability parts of Delphi. His decisions are in light of the work of numerous different specialists, such as Dalkey and Helmer (1963), Campbell (1966), Pfeiffer (1968), Dalkey (1969), Farquhar (1970), Martino (1970), Gustafson *et al.* (1973), Hill and Fowles (1975) and Parente *et al.* (1984)	Expert opinions are generally used when there is a little or no past data available or in the condition when organisations want to maintain the secrecy of their end-product, process, services or technology before introducing time
2. Trend analysis Trend impact analysis Trend extrapolation Growth curve fitting Precursor analysis Long wave analysis	Trend analysis includes prediction through the continuation of quantitative past information into what is to come. It includes a broader term that incorporates monetary models and methods of forecasting, for example, exponential smoothing, regression, ARINA model, Box-Jenkins' and growth curve fitting (Levary and Han, 1995). Every technology has a life cycle made out of a few particular stages. This life cycle ordinarily consists of an introductory stage, a development, maturity and the last declining stage. In light of parameter estimation of life cycle assessment, growth curve can be determined. The growth curve technique of forecasting is useful in evaluating the maximum furthest reaches the level of growth of technology or a decrease in the phase of life cycle. It is used for determining and useful in anticipating innovation life cycle stages, i.e. when innovation will achieve a specific stage in the life cycle (Kivikunnas, 1998). One of the most common growth curve technique of forecasting is Fisher–Pry analysis. It is utilised for anticipating dissemination of innovations and applies mainly to those innovations which do not oblige significant behavioural changes. One can use nominal group technique (NGT) as a substitute for Fisher–Pry analysis which is given by Delbecq and Van de (1970)	Trend analysis methods are used to benchmark when the organisation continuously produces the technologies, product and services. It is a quantitative approach which requires enough past data for analysis and is generally used by the textile, automobile and electronics industry
3. Monitoring and intelligence methods Bibliometrics (patent analysis, text	Monitoring variations, technology watch and environmental scanning are more suitable for rolling out mindful of changes that could affect the infiltration or acknowledgement of the innovations in the commercial centre (Phillips *et al.*, 2007). Madnick and Woon (2009)	Monitoring and intelligence methods are used where researchers or industry want to explore new areas of

(*continued*)

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

Table I.

TF family and TF methods	Brief description	Conditions for using TF methods
mining, research profiling) Monitoring (technology watch, environmental scanning)	states "environment examining/scanning can be considered as the focal point to prospects research. However, the results are generally broad, making it impossible to bolster a particular choice". Monitoring system goal is to discover early signs of conceivably imperative innovations to pick up however much lead time could be expected (Van den Hende *et al.*, 2007). Resource accessibility is one of the scoping issues connected with these techniques since various examining methodologies oblige the utilisation of specialists (Firat *et al.*, 2008; Huang *et al.*, 2014)	development and innovations. It is mostly used by the R&D department for R&D purposes and not by the managerial department
4. Statistical methods Risk analysis demographics Correlation and regression analysis ANOVA Cross-impact analysis	A statistical model is reliant on a progression of observations, and it outlines the pattern of different variables that are of interest by establishing the relationship between them. Descriptive models utilised as a part of forecasting are usually quantitative; however qualitative models are also utilised. Descriptive phenomena are single event occasions, and like this, they are troublesome to model. In this way, the use of a statistical model requires an intensive comprehension – stretching the anticipating procedure (Firat *et al.*, 2008). In the family of statistical methods, correlation and regression analysis are the most commonly used methods. Correlation and regression analysis figure the development of the pattern of the advancement of the innovation which is similar to those of existing technology. Utilisation of this technique presupposes the data concerning the development pattern of the current technologies which are accessible	Statistical methods are used where researchers or managers want to know the significance level of the different parameters of product, process, service or technology used by the organisation
5. Modelling and simulation Cross-impact Analysis Agent modelling Input–output analysis Diffusion modelling Systems simulation (KSIM, system dynamics) Sustainability analysis Scenario-simulation (gaming; interactive scenarios) Technological substitution Economic base modelling Causal models Complex adaptive system	Modelling obliges a decent comprehension of cooperation between forecast and the hidden variables or determinants. Models can help to display the future conduct of complex frameworks basically by secluding vital framework aspects from the unessential point of interest. (Firat *et al.*, 2008) System simulation is one of the famous techniques in this crew. The significant advantage of systems simulation is that "it permits clients to hunt down the best ways to deal with a problem, confronting a test, or tackling an issue, without the danger or cost of unreasonable mix-ups" (Madnick and Woon, 2009). Given this advantage, it is conceivable to envision a system which contains all (or the same number of as sensibly conceivable) technologies and it helps to figure out which innovation has the most elevated future worth (De Holan *et al.*, 2004). Despite the fact that it is conceivable to envision the model hypothetically, the handy parts of actualising such a model would be overwhelming because the exactness of that system simulation depends on the authenticity of the data and on the way in which relationships are communicated in the simulated model	Most of the industrial problems are complex which are sometimes impossible to analyse in the real environment because of the cost factor. In that case, modelling and simulations tools are used by considering the real condition of the business environment

(*continued*)

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

TF family and TF methods	Brief description	Conditions for using TF methods
modelling (CAS) (Chaos) Life cycle analysis		
6. Scenarios Scenarios (scenario management, scenarios with consistency checks) Field anomaly Relaxation method (FAR) Scenario-simulation (interactive scenarios, gaming)	Scenario composing proposes diverse originations of future innovation. Every origination of the innovation is based on a very much characterised arrangement of presumptions. A scene shows the option qualities of the future innovation with every option being given specific suppositions and conditions. The forecaster assesses the legitimacy of the presumptions. The after effects of this assessment are utilised to focus the situation destined to happen (Levary and Han, 1995)	Scenarios are generally used by researchers and admiration to predict the possible as well as preferable future
7. Value/economic/decision methods Action analysis Economic base modelling Cost-benefit analysis Relevance trees (futures wheel) Decision analysis (utility analyses)	Relevance tree approach is the most prominent technique in this family. It is a normative way to deal with TF (Esch, 1972). The different objectives of innovation are separated and arranged in a tree-like arrangement. In this arrangement, the various levelled of the objective of the proposed innovation is formed. Evaluation of the probabilities of accomplishing the objective at the different levels of technological advancement must be done. The probabilities are then utilised to forecast the likelihood of accomplishing the objectives and targets of the innovation which is proposed (Levary and Han, 1995)	Value/economic/decision methods are generally quantitative analysis which is used by the administration in decision making only by considering the financial profit or cost benefit
8. Descriptive and matrices method Back casting Requirements analysis Checklist for impact identification Analogies Innovation system modelling Mitigation analysis Institutional analysis Road mapping Morphological analysis multiple perspectives assessment Social impact assessment Organisational analysis	A developing action in this class is road mapping, which undertakes major technological tasks in manufacturing and product design together for effectively attractive breakthroughs. Roadmaps run an innovation from two to ten years ahead. In a broader term, an innovation roadmap gives a vision or perspective of tomorrow's technologies which is accessible to users. Previously, the institutional champions for road mapping were formed military associations. Now, they have been other expansive partnerships and industry affiliations (Coates *et al.*, 2001) Analogies are likewise prevalent descriptive methods. The utilisation of analogies in forecasting includes an orderly correlation of the innovation to be estimated with some prior innovation that is comparable in all or most vital regards. But as indicated by Firat *et al.* (2008) that "In any case, there is no assurance that people of today will make same choices as individuals did in the prior circumstance. Consequently, the forecast is at most plausible, never certain"	The researchers generally use descriptive and matrices method during the initial design phase of an innovation where researcher try to find the all possible alternative solutions

Table I.

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

Observation 3: various TF tools/methodologies may be utilised towards availing vital benefits from TF outcomes regarding valuable predictions and recommendations:

- Expert opinion – it may be utilised for forecasting and innovative improvements by using expertise and experience of specialists in their respective fields. Delphi has been reported as the most prevalent tool to transform subjective opinions into objective recommendations.
- Trend analysis – it utilises the continuation of quantitative past data to predict what is to come and is useful in evaluating the utmost level of technology growth that is followed by a decrease in various phases of technology life cycle and forecasting when innovation is expected to achieve a specific life cycle stage.
- Monitoring and intelligence methods – it may be utilised to predict the lead time and improvements' short-term impacts on acknowledgement and adoption of innovations commercially, and this family include tools such as monitoring variations, environmental scanning, technology watch and resource accessibility.
- Statistical methods – by analysing statistical data to delineate a behavioural pattern of interactions among various variables associated, the most prominent strategies are correlation analysis (for developing patterns to forecast the timing of innovation advancement utilising existing technologies) and bibliometrics (that includes research profiling, patent analysis and text mining).
- Modelling simulation – it may be employed to predict the behaviour of various variables/actors to display the future conduct of complex frameworks helpful in dealing with comprehension of conflict/cooperation between forecast and hidden variables/determinants.
- Scenarios – they may be utilised to propose diverse originations of future innovation based upon well-categorised presumptions' arrangement to assess options' qualities towards focusing on the situation most intended to occur.
- Values/decision/economic methods – these relevance tree approach, which is the most used technique of this family, is a normative way to evaluate probabilities of accomplishing the goals arranged in the form of tree-like levelled structure to predict the likelihood of achieving well-defined objectives of innovation.
- Descriptive and matrices methods – these include innovation road mapping (that takes care of key technological items to predict and provide guidelines for innovations' or products' generations) and analogies (that utilises orderly correlation to estimate with some prior innovation assuming almost similar conditions/ situations).

Furthermore, it is important to mention here that the TF methods are also divided into two categories, i.e. "exploratory" or "normalising". This paper also partitions TF systems into exploratory, normative and the blend of two categories, as per Technology Futures Analysis (2004) Methods Working Group, and the same is also shown in Table II.

Observation 4: all TF tools/methodologies may be categorised into three groups: exploratory (to anticipate the technological state that is likely to be in the future), normalising (to predict and estimate about what technologies need to be or should be conceived at some near future time), and normative/exploratory.

4. Relationships among TF methods

In this segment of paper, we endeavour to evaluate the relationship between exploratory and normative methods that are given in the past literature and try to recognise the linkages among the methodologies. Some TF strategies are firmly utilising both exploratory and normative methods to anticipate the technological development or changes. Moreover, the choice of fitting TF methods relies upon the nature of the technologies (Mishra *et al.*, 2002). Along these lines, it obliges experience and skill in different TF procedures to choose the suitable estimating models. In fact, this section classifies TF procedures as indicated by exploratory and normative methodologies. Table III demonstrates a lattice of TF methods by sorting strategies and technological characteristics inside every cell, and TF methods are recorded in the decreasing order of their utilisation frequency.

There are no fixed philosophies in technology anticipation or forecasting, and a mix of distinctive methodologies and techniques are generally used to enhance the precision of estimation. Methods hybrids are better than a solitary method (Schnaars, 1989). A blend of different strategies empowers forecasters to examine different points of view (authoritative, innovation, individual, social and natural) (Daim *et al.*, 2006). The specialists in TF contend that the multifaceted nature technologies and quick social change obliged the requirement for developing tools that combine exploratory methods with normative systems (Coates *et al.*, 2001). Over the most recent decades, particularly after the rise of ICT, a percentage of the distinctive methodologies has been developed and utilised by diverse scientists brushing with numerous different tools. Figure 1 exhibits the heritage of TF strategies and also there are some limitations of TF that are systematically analysed and compiled in a tabular form.

Term	Definition	Characteristics
1. Exploratory	The endeavours to anticipate the technological state that is likely to be in the future	Undertaking foreseen results Recommend distinct options for the proposed allocations
2. Normative	The announcement of what technologies need to be or should be conceived at some near future time REF	Excessively complex and numerically complicated Important when objectives are critical and specific Acknowledgement of monetary possibilities Attention to limitations (normal assets, organisation assets and so on) Acknowledgement of a technological potential
3. Normative/ exploratory	Can be used in two different approaches	They possess the characteristics of both Normative and exploratory

Sources: Compiled by the authors according to Roberts (1969), Fisher and Pry (1971), Twiss (1984), Martino (2003) and Porter (2010)

Table II. Categories of TF methods

Discontinuous		
Bibliometrics (patent analysis, text mining, research profiling)	Trend Impact Analysis	Decision analysis
Cross-impact analysis	Delphi	Morphological analysis
Input–output analysis	Scenario planning	AHP
Diffusion modelling		Relevance tree
Continuous		
Trend extrapolation	Road mapping	Morphological analysis
Growth curve fitting	Trend impact analysis	AHP
Precursor analysis	Delphi	Relevance tree
Long wave analysis	Scenario planning	
System dynamics		
Exploratory	Exploratory/Normative	Normative

Source: Adopted from Daim *et al.* (2006)

Table III. A lattice of TF methods

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

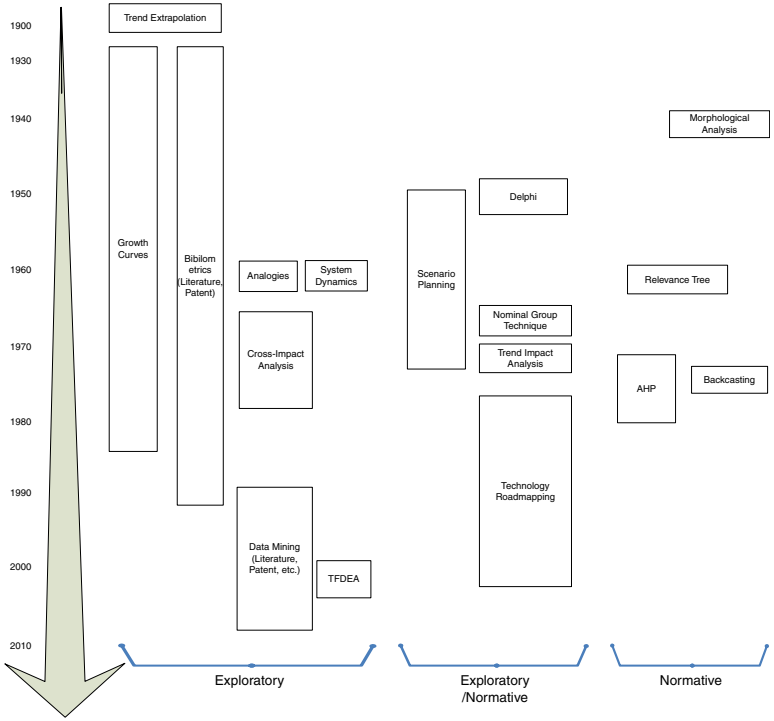


Figure 1.
The heritage of TF strategies

Source: Adopted from Daim *et al.* (2013)

4.1 Systematic analysis of limitations of TF

A variety of techniques have been developed for technological forecasting. As in all other forecasting methodologies, the most effective ones are based on careful analyses of experience combined with the insights of competent and imaginative people. Each method has its limitations, and we have tried to analyse the limitations of TF systematically. Table IV tabulates the inadequacies of TF techniques.

From this section, we may conclude the following observation.

Observation 5: combination of various tools/methodologies may be used towards effective and efficient forecasting of technology with ample utilisation of information and communication technology for identifying the optimum solution.

5. Technology assessment

The TA concept is more than a quarter century old, and the plethora of research articles is already available in this subject area. However, it is still not clear in the mind of researchers other than its experts. TA is neither simple forecast nor prediction research, or neither social effect analysis nor absolutely system analysis (Finsterbusch, 1980). It deals with the entire system that analyses the effects and the causes (i.e. immediate or

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

Inadequacies of technology forecasting	
Limitation of quantitative techniques	Exactness, stability and dependability are adversely affected by long-term forecast
	Adaptability to present rate of progress in innovation is troublesome
	Complex models hard to fathom and practice
	The problem in fitting unruly data into the mathematical straight jacket
	Comparative static and reorientation is difficult
	Real-world feedback is difficult to carter
	Lack of adequate past data
Limitation of qualitative techniques (exploratory)	Definition and determination of professional and specialist is troublesome
	Individual inclination goes into a subjective appraisal
	Assigning obligation and responsibility to forecaster is troublesome
	The validity of assumption in situation improvement tends to questionable with time
	Social and political elements are disregarded generally
	Inconsistent assessments and inability to recognise probabilities concerning distinctive timelines in cross-impact
	Number of rounds, nature of expert, precision and dependability has inborn weakness in Delphi
	Sharps interruption of patterns and sudden occasions are evaded normally
	The timescale is either optimistic or pessimistic
	The timescale is not coordinated with innovation
	Forecast confined to inside perspective of the association
	Impingement of other innovations are not considered
Limitation of qualitative techniques (normative)	Dynamics of framework get disregarded in the minuscule vision of innovation
	Maybe connected with passage vision without due respects to the environment
	Technology at the base of the tree may be biased
	Technology choice for the target may fall into a generic category
	Difference between normative and exploratory forecast is often blurred
Difficulties in translating TF into Technology implementation	Commitment is lacking
	Attention to the flexibility of TF in technology implementation is inadequate
	Attention to dynamic tension between exploitative and exploratory activities in an on-going operation. It is thoughtless
	Inadequate to simultaneous rather than sequential development
	Selection of modular technology is neglected
	No attention to dynamics of the impact of technology on organisation
	No critical review and midcourse correction
	Benchmarking and enhance technological performance measurement due to TT is lacking
	Selection of cross-functional team consisting of an expert of TF, TT, engineering, etc., is neglected
	Attention to in-house R&D is inadequate
	TF, TT and implementation are relegated to separate group that experience looping, design looping, and concept crossing are given lesser importance

(*continued*)

Table IV.
Inadequacies of technology forecasting techniques

Inadequacies of technology forecasting	
Human-related problem with forecasters	Selection of the team of the forecaster is not correct
	No-real mean to define an expert who is making a forecast
	The forecasters are not made accountable
	Avoiding criticism, feedback and reorientation for the initial forecast
	Tendencies to prevent information gathering and scanning before the forecast
	Tendencies to resort to immediate gains by incremental developments rather than making radical development as long-term planning
	Individual bias is associated with forecasts and fixing weightings
	Expert block looking at forecast only through one's expertise without considering to parallel developments
	Creative blocks in vision due to social insecurity, fear of consequences, initial and perception factors
	There is a tendency to concentrate on specific configuration rather than considering extrapolating aggregate figure of merit.
	Existence an intellectual rift between forecaster
The problem in selection of technique	There is a mismatch of a technique for the technology to be forecasted
	Overlap of the forecast for continuity is disregarded
	The technique is forced to perceived certain technology
	Validation of technique by another is not carried out to reduce subjectivity
	The technique does not incorporate performance measures

Table IV.

after-effects, sometimes communicated as first-, second- and higher order effects). It finds out whether the effects have been arranged or proposed in true circumstances that are portrayed into technological improvement; furthermore, it explains the advantages or unfavourable nature of the consequences and typically forces investments on a few sections of society that are beneficial.

In an industrialised developed economy, TA may been viewed as an innovation strategy. In developing the economy, TA is viewed as another device for social administration of innovation. TA is relied upon to help in selecting proper advancements for improvement. TA has various views in different economies (Flynn and Bellaby, 2007).

TA generally manages vast and unruly systems that cannot undoubtedly be displayed thoroughly as per the fundamentals of the framework and cannot definitively profit by a great quantitative control of data. Besides, it includes the development and arrangement of choices for economist and decision makers. TA considers the social effects, regarding instability and its consequences, a complete endeavour to study and explore the real issue areas.

5.1 The need for TA

TF methodologies will be important, but not adequate, towards innovation and sustainable development. Thus, we will apply an amplified thought of advancement, which incorporates specialised angle furthermore social and institutional perspectives (Fleischer and Grunwald, 2008). The finding of the general inner conflict of innovation concerning manageable improvement can be changed into a solicitation for sufficient moulding of innovation: innovation and its societal surroundings ought to be produced further and shaped in a manner that its positive outcomes help in economic advancement and negative ones be avoided or minimised.

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

This is the inspiration for TA: to empower society and to reap the advantages of a particular innovation without running into circumstances of danger for the supportable environment.

5.2 Tools of TA

In writing, there is no specific procedure concerning TA to settle upon. Maybe it is contended that every innovation requires the scope to apply proper methods to a particular circumstance, as opposed to taking after a methodology that is generic (Coates, 1976; Wood, 1997). Also, there is a conceivably different image of TA and its significance that is generally contingent upon the nation and society (Chen, 1979) (Table V).

From this section, we may conclude the following observation.

Observation 6: TA has been recognised as a less revealed and utilised technique to manage vast and unruly systems for assessing long-term consequences of technology to carry out the social administration of innovation for empowering society to reap the advantages of a particular innovation towards sustainable creation/development of the nation.

6. Discussion of findings

It is a fact that the pertinence of TF and TA in today's world is more noteworthy than it was in the past. A lot of assets are piped into innovation systems landed by utilising TFA. In this way, TF and TA impressively have been turn out to be more result-oriented, effective and successful in the present scenarios. The technological change and innovation are both drivers that help in the setting financial, social and political prosperity. Envisioning and comprehending the course of innovative change is problematic for decision makers in both businesses and government organisations. Some techniques are being utilised for TF and TA. Minimal orderly emphasis has been given to the calculated improvement of the TF and TA field overall. With a specific end goal to make TF and TA more powerful, one of the issues that should be tended to is a legitimate choice of the method while accounting its constraints. A conceptual model of TF and TA has been presented in Figure 2, after consulting the experts and reviewing the research text available in various research journals of national and international repute.

6.1 Implications of the research

Examinations of rising advances and their suggestions educate significant choices beginning from the worldwide level to the individual association.

For instance, large organisations need TF and TA, in its different section:

- to prioritise research and development;
- in NPD; and
- in fabricating key decisions on innovation authorising or finding the critical factor that influences the technologies, etc.

Small companies conjointly rely on innovative advancement for their presence. In these companies, TF ways are utilised to gauge appropriation or dissemination of innovation. However, with the exertion to plan financially reasonable guides for technological advancement by organisations, the TF and TA field incorporates or diffuse social estimates too. Apart from this, the study can be used in practice for identifying the method that can analyse the economic, commercial and sustainable outputs. This study can also be used as a teaching material of TF and TA methodologies. Society as a whole is impacted by the TA methodologies as its methodologies help in sustainable development, which influences the public policies. A framework presenting learning, action/s, actor/s and outcomes as deduced from findings of the paper has been put forward in Table VI.

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

S. No.	Methods of Technology Assessment	Description of Technology Assessment Methods
1	Case study/analysis	Case study analysis includes the "depth examination of past episode to create clarifications that may be generalised to different occasions". Mostly case study analysis is a profundity that supports the different proof of relationships that are causal (Gerring, 2004; George and Bennett, 2005) In this method of technological assessment, there is a case analysis of comparable technologies which gives information about the potential issues that may emerge, nonetheless, where innovation is novel and where similar circumstances are not accessible case study analysis is constrained (Raven *et al.*, 2009)
2	Cost-benefit analysis	A monetary analysis that compares the expenses and advantages, typically evaluated in money-related terms, for situations with and without activity or action (Tran and Daim, 2008)
3	Cost-effectiveness analysis	An economic analysis that gives info about the expense viability of option for accomplishing a goal, and is utilised to find the option with the minimal direct money-related expense (Tran and Daim, 2008)
4	Delphi technique	In the Delphi technique, there is a gathering of individuals (frequently a panel of applicable professionals or experts) who make a judgment and express their suppositions on the given issue. These suppositions are evoked using various rounds of questionnaire or surveys (alluded to iteration) (Rowe and Wright, 1999) with feedback between each round to illuminate the conclusions with the help of their colleagues. This procedure of correspondence is intended to permit a group of people to survey their suppositions in light of the learning of different members (Jain *et al.*, 1993, p. 204). After a few rounds, the last agreement is taken into account the measurable mean/median of the members' conclusions (Rowe and Wright, 1999; Jain *et al.*, 1993). The key components of the Delphi system are the emphasis, anonymity (accomplished with the utilisation of questionnaire) and controlled feedback and also a factual conglomeration of a gathering reaction (Haque *et al.*, 2013)
5	Choice modelling	An experimental strategy utilised as a part of financial matters to edge trade-offs between diverse alternatives. The strategy assesses the estimation of alternatives by uncovering that how much respondents has a will to trade them. Through surveys, data are generally accumulated (Tran and Daim, 2008)
6	Focus groups	A group meeting strategy where a facilitator offers a question to start the conversation among members (Kerr and Tindale, 2011)
7	Impact pathway analysis	It used to anticipate the pathway of effects coming about because of some certain action. This system endeavour to insight into the immediate and backhanded effects of activities and interaction. Otherwise called change pathways (Tran and Daim, 2008)
8	Interviews	A strategy for essential information gathering that comprises top to bottom addressing. The interview may vary as indicated by the sort of informant, the kind of medium (phone or *vis-à-vis*, individual or gathering), and the addressing (organised, semi-organised and unstructured)
9	Input–output analysis	It examines the connections and relationships of output productivity through an examination of the flow of various resources. It is the analysis of the inputs to industry, exchanges between parts, family unit utilisation and the yields of products delivered
10	Life cycle assessment (LCA)	LCA is a broadly perceived methodology for evaluating the effects of a product or a process over its whole life cycle, and it can apply on any process (Stewart, 1999, 2001; Stewart and Petrie, 2006; Norgate *et al.*, 2007). It has also been utilised as a technique in the assessments of the environment which are for a long period. It is already demonstrated that a frameworks approach (as utilised as a part of LCA) is important in finding open doors for advancement, budgetary funds and development of environmental friendly condition (Bossel, 1999). Over a few years, there is an attempt to include both social and financial components into LCA (Brent and Labuschagne, 2006; Jeswani *et al.*, 2010)
11	Social and regional profiling	A procedure to gather important essential and auxiliary information around a certain section of the community. It helps in profiling an elaborate depiction of the group, environment and economy of a district and gives info about qualities, needs and patterns (Howitt, 1989). It is an evaluation of the present condition of a group or social community which also includes trends. Knowing the certain community helps with envisioning how individuals may react to change (Tran and Daim, 2008). Understanding community includes an investigation of their connections and systems and the qualities that may shape their state of mind and practices. Profiling incorporates examination of demographic patterns, populace attributes, ethnicity and society, the neighbourhood economy, work business sector, area utilisation, social and political association, family and group association, well-being, nourishment, illness, group base and administrations (lodging, well-being, childcare and so forth), group "needs" and wanted fates and the ability to address these issues (Slootweg *et al.*, 2001)
12	Scenario analysis	It is a methodology to acknowledge the change in diverse conceivable future circumstances. Scenario analysis can help to get ready for impromptu planning. On the off chance that directed with groups, scenario analysis can help to educate the general population of risk and helps in managing their goals (Jolly, 2008)
13	Social mapping	A procedure for distinguishing and tracking the importance and qualities credited to a certain area by social communities (Assefa and Frostell, 2007)
14	Consensus conferences	It includes the get together of a chose group of natives to find out around a given innovation and inquiry the technologists on their worries. The procedure ordinarily includes various workshops; with nearly consider 15 natives are taking an interest. These conferences are especially mainstream among European innovation evaluation advocates (Kleinman *et al.*, 2007), it has also been shown some accomplishment inside of an Australia too Einsiedel and Eastlick (2000)
15	Stakeholder analysis	It comprises of selecting of partners; investigation of their fundamental mentality and inspirations, a determination of which partners are most critical, a comprehension of their systems and connections and the improvement and execution of an engagement arrangement (Dewulf and Van Langenhove, 2005). A partner/ stakeholder is any individual who can influence, or is influenced by, an organisation's activities. Focussed on consultation (or centred arrangements) should be created for every partner,

(*continued*)

Table V.
List of methods of technology assessment through different categories

(*continued*)

Table V.

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

S. No.	Methods of Technology Assessment	Description of Technology Assessment Methods
		particularly for a vulnerable group. Partners have changing degrees of force, authenticity and enthusiasm for an issue or a venture. Partners additionally incorporate individuals inside of organisations that may be imperative to the arranging, improvement and execution of the action (Hoffmann *et al.*, 2004)
16	Strategic and regional assessments	It may be embraced during or before establishing a new kind of organisation. The upside of such methodologies is that they: encourage the early recognisable proof and determination of potential issues when there is the adaptability to roll out improvements, give a chance to longitudinal and comparable examination(Stewart and Petrie, 2006). It helps in more successfully recognise existing and potential aggregate effects, might expressly connect evaluation to provincial arranging and reporting and can set up the standard and territorial data sets that help the improvement of area wise checking endeavours. It can be the most proper type of evaluation for different regions including multiple partners and have large scale activities (Evans *et al.*, 2007)
17	Sustainability assessment and metrics	In sustainability assessment, there is a term that covers a large scope of approaches with no specific set of methodology. It inspects an innovation, technologies or processes in the light of economic advancement criteria or measurements, an endeavour to recognise changes or effects on ecological, social and financial bases (Khurana *et al.*, 2014). Physical sciences methodologies have a tendency to command sustainability assessment, particularly as to innovation appraisal where numerical methodologies are frequently looked to give some apparent elevated amount of assurance than subjective depictions (Jischa, 1998; Azapagic, 2004; Hoffmann *et al.*, 2004; Dewulf and Van Langenhove, 2005; Pinter, GRI, 2006)
18	Trend analysis	The accumulation and examination of verifiable and contemporary information to illuminate the forecast of future (Tran and Daim, 2008)

Table V.

6.2 *Unique contributions of the research*

TF and TA have been the developing area of management studies for decades (Nikolopoulos *et al.*, 2015). Some studies on TF and TA have been reported in the past and have helped in the development and advancement of this field. However, there is no endeavour made to introduce a compiled review of the methodologies and its tools that have been referred to TF and TA (Altuntas *et al.*, 2015). In this study, we try to fill that void. An intensive audit of the TF and TA from the past studies is led to the exploration of different tools and methodologies towards providing the following observations.

- Experts' opinions are important to transform subjective opinions into objective recommendations towards carrying out forecasting and innovative improvements.
- To predict future, trend analysis utilises continuation of quantitative past data, whereas, monitoring and intelligence methods are being utilised to predict short-term impacts.
- Statistical data and its analysis are used to delineate a behavioural pattern of interactions among various variables associated; also, modelling simulation can be found useful to predict the behaviour of various variables/actors to deal with conflict/cooperation between forecast and hidden determinants.
- Scenarios may be utilised to propose different originations of future innovation-based presumptions' arrangement; however, the relevance tree approach (one of values/decision/economic methods) uses probabilities of accomplishing the goals. Also, descriptive and matrix methods include innovation road mapping and analogies.
- TF tools/methodologies may be categorised into three groups: exploratory, normalising and normative/exploratory; however, a combination of various tools/methodologies can be a good option towards effective and efficient TF with ample ICT utilisation.
- Furthermore, TA has been observed as a less revealed and utilised technique to manage vast and unruly systems.

The paper may be helpful in selecting the appropriate method of TF and TA for a specific situation. An exhaustive presentation of the methodologies/tools has been introduced.

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

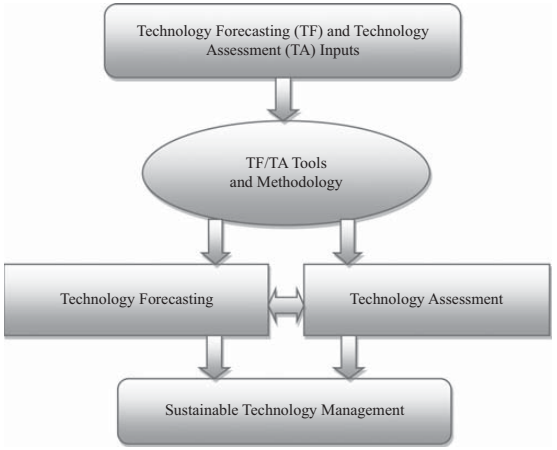


Figure 2.
A conceptual model of technology forecasting and technology assessment

7. Conclusion

We should refrain from considering forecasting and assessment as a solitary action with "one size fits all" philosophy. This study concludes that the quest for a single strategy for doing forecast and assessment is a misconception because this neglects to perceive that forecast and assessment oblige a suitable blend of strategies and methods drawn from a variety of fields. The proper blend of techniques to be utilised, the variables to be chosen, the units of estimation to be utilised and the grouping of analysis must remain a function of the problem in hand and the experience and inventive, creative ability of the investigator(s).

This study is an in-depth literature review of TF and TA using the paper mainly available from 1960 to 2015. This study is more general in application, for future research, one can study more industry specific or problem specific. Furthermore, one can also include the different factors that can lead to the selection of a particular methodology.

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

Observation	Learning	Action/s	Actor/s	Outcome/s
Observation 1: TF and TA needs to be explored regarding tools/ methodologies	Research efforts need to be fostered in TF and TA area	Initiation and actualisation of research programs at fast track	Researchers, research organisations, academicians, universities	Research findings
Observation 2: not ample literature available on TF and TA tools/ methodologies relevant to new product development (NPD)	The literature on TF and TA to be made available relevant to new product development (NPD)	Organising conferences, seminars, workshops, etc.	Researchers, research organisations, academicians, universities	Developing countries' relevant literature
Observation 3(a): expert opinion may be utilised for forecasting and innovative improvements to transform subjective opinions into objective recommendations	Utilising expertise and experience of specialists has been found vital to transforming subjective opinions into objective recommendations	Identifying specialists/ experts of relevance from diverse disciplines; contacting them and arranging for the appropriate conducive environment for interactions (direct/ indirect)	Specialists/experts, organising crew	Valuable recommendations
Observation 3(b): trend analysis utilises continuation of quantitative past data to predict future	Quantitative past data may be utilised towards TF	Identifying relevant past data; sorting and correcting data to make it usable for trend analysis; arranging for appropriate software and skilled professional; carrying out trend analysis	Data mining organisation/experts, IT professional/s, Trend analyst/s	Utmost level of technology growth that is followed by a decrease in various phases of technology life cycle and forecasting when innovation is expected to achieve a specific life cycle stage
Observation 3(c): monitoring and intelligence methods may be utilised to predict short-term impacts and tools are monitoring variations, environmental scanning, technology watch and resource accessibility	Monitoring and intelligence methods may be useful to forecast short-term gains in commercial terms	Keeping a close watch on technological changes, developments and innovations nationally and globally and their impacts on various stakeholders/ nation/country's counterpart	Monitoring and intelligence methods' experts	Lead time and improvements' short-term impacts on acknowledgement and adoption of innovations commercially
Observation 3(d): by analysing statistical data to delineate a behavioural pattern of interactions among various variables associated, and the most used strategies are correlation analysis and bibliometrics	Statistical data availability may be followed by statistical analysis for authentic forecasts and predictions	Making suitable data available, apply statistical tests for its analysis towards making forecasts	Statistical analysis experts/specialists, IT expert/s, TF expert/s	Behavioural pattern of interactions among various variables associated; most acceptable recommendations based upon statistical forecasts
Observation 3(e): modelling simulation may be utilised to predict the behaviour of various variables/actors helpful in dealing with conflict/ cooperation between forecast and hidden determinants	Simulation and modelling may reveal hidden determinants' behaviour that further makes decision making effective and reliable	Identifying variables/ actors, arranging for suitable data and modelling/simulation software, carrying out simulation/modelling, analysing results	Firms/individuals making suitable data availability, simulation expert/s, IT expert/s	The behaviour of various variables/ actors to display future conduct of complex frameworks
Observation 3(f): scenarios may be utilised to propose diverse originations of future innovation based upon well-categorised presumptions' arrangement to assess options' qualities towards focusing on a situation most intended to occur	Diverse originations of future innovation may be predicted by using scenario technique	Identifying/finalising presumptions, identifying options and their qualities, carrying out the assessment of options for a situation most likely to happen	Scenario experts	Assessment of qualities of alternatives/options
Observation 3(g): value/ decision/economic methods- relevance tree approach, which is the most used technique of values/decision/economic methods' family, showing probabilities of accomplishing the goals arranged in the form of tree-like structure to predict the likelihood of achieving objectives of innovation	Probabilities of accomplishment of various objectives of innovation/s may be evaluated using a relevance tree approach which may further facilitate to make recommendations/ decisions	Defining the problem, arranging probabilities of accomplishing the goals arranged in the form of tree-like structure, analysing the probabilities' structure, making predictions	Relevance tree expert/s, technology forecasting specialist/s, mathematician (expert in mathematical operations on probabilities)	Probabilities of goals accomplished in the form of levelled structure, forecast exhibiting a likelihood of achieving well-defined objectives of innovation
Observation 3(h): descriptive and matrices methods include innovation road mapping and analogies	Long-term TF and planning may be done by using this category of TF	Gathering relevant facts and figures, preparing innovation roadmaps/analogies towards log terms predictions	Highly experienced and skilled TF experts	Technology/ innovation roadmap/ s, long-term guidelines for innovations' or products' generations, analogies
Observation 4: all TF tools/ methodologies may be categorised into three groups: exploratory (to anticipate the technological state that is likely to be in the future), normalising (to predict and estimate about what need to be conceivable at some future time) and normative/ exploratory	A clear understanding of three categories of TF tools need to be developed, every category has its advantages and limitations	Developing a clear understanding of TF tools and their respective category, developing a clear understanding of TF tools' relative advantages, disadvantages and applications	TF expert/s	Better understanding and facilitation of choice making of TF tools from three categories depending upon nature of problem/ situation/technology
Observation 5: combination of various tools/methodologies may be used towards effective and efficient TF with ample ICT utilisation	Different TF tools may be combined towards achieving hybrid methodology o overcome limitation/s of tools when used individually	Evaluating various combinations of TA tools regarding their advantages and limitations	Specialists/experts	Hybrid TF methodology for optimum solution
Observation 6: TA has been recognised as less revealed and utilised technique to manage vast and unruly systems	TA tools need to be appropriately analysed and utilised	Assessing long-term consequences of technology, social administration of innovation for empowering society	Government, international organisations, TA specialists/experts, data mining experts/firms; Stakeholders, general public	Sustainable creation/ development of the nation

Table VI.
Learning and actions

(*continued*)

Table VI.

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

References

Adebanjo, D. and Mann, R. (2000), "Identifying problems in forecasting consumer demand in the fast moving consumer goods sector", *Benchmarking: An International Journal*, Vol. 7 No. 3, pp. 223-230.

Altuntas, S., Dereli, T. and Kusiak, A. (2015), "Forecasting technology success based on patent data", *Technological Forecasting and Social Change*, Vol. 96, pp. 202-214.

Assefa, G. and Frostell, B. (2007), "Social sustainability and social acceptance in technology assessment: a case study of energy technologies", *Technology in Society*, Vol. 29 No. 1, pp. 63-78.

Azapagic, A. (2004), "Developing a framework for sustainable development indicators for the mining and minerals industry", *Journal of Cleaner Production*, Vol. 12 No. 6, pp. 639-662.

Bijker, W.E. (1992), "The social construction of fluorescent lighting, or how an artefact was invented in its diffusion stage", *Shaping Technology/Building Society*, pp. 75-102.

Bossel, H. (1999), *Indicators for Sustainable Development: Theory, Method, Applications*, International Institute for Sustainable Development, Winnipeg, p. 138.

Brent, A. and Labuschagne, C. (2006), "Social indicators for sustainable project and technology life", *The International Journal of Life Cycle Assessment*, Vol. 11 No. 1, pp. 3-15.

Bush, V. (1990), *Science: The Endless Frontier*, National Science Foundation (NSF 90-8), Washington, DC, (Orig. ed. 1945).

Campbell, R.M. (1966), *A Methodological Study of the Utilization of Experts in Business Forecasting*, University Microflims, UCLA.

Chen, K. (1979), "International perspectives on technology assessment", *Technological Forecasting and Social Change*, Vol. 13 No. 3, pp. 213-233.

Coates, J. (1976), "The role of formal models in technology assessment", *Technological Forecasting and Social Change*, Vol. 9 Nos 1-2, pp. 139-190.

Coates, V., Farooque, M., Klavans, R., Lapid, K., Linstone, H.A., Pistorius, C. and Porter, A.L. (2001), "On the future of technological forecasting", *Technological Forecasting and Social Change*, Vol. 67 No. 1, pp. 1-17.

Compagna, D. and Kohlbacher, F. (2015), "The limits of participatory technology development: the case of service robots in care facilities for older people", *Technological Forecasting and Social Change*, Vol. 93, pp. 19-31.

Daim, T., Oliver, T. and Kim, J. (2013), *Research and Technology Management in the Electricity Industry*, Springer, London.

Daim, T., Rueda, G., Martin, H. and Gerdsri, P. (2006), "Forecasting emerging technologies: use of bibliometric and patent analysis", *Technological Forecasting and Social Change*, Vol. 73 No. 8, pp. 981-1012.

Daim, T.U. and Dash, P. (2011), "Roadmapping the convergence of technologies for services over broadband: a benchmarking effort", *Benchmarking: An International Journal*, Vol. 18 No. 5, pp. 668-693.

Dalkey, N. (1969), "An experimental study of group opinion", *Futures*, Vol. 1 No. 5, pp. 408-426.

Dalkey, N. and Helmer, O. (1963), "An experimental application of the DELPHI method to the use of experts", *Management Science*, Vol. 9 No. 3, pp. 458-467.

De Holan, M., Pablo, N.P. and Lawrence, T.B. (2004), "Managing organizational forgetting", *MIT Sloan Management Review*, Vol. 45 No. 2, pp. 45-51.

Delbecq, A. and Van de, A. (1970), "Nominal group techniques for involving clients and resource experts in program planning", *Academy of Management Proceedings*, Vol. 1970 No. 1, pp. 208-227.

Dewulf, J. and Van Langenhove, H. (2005), "Integrating industrial ecology principles into a set of environmental sustainability indicators for technology assessment", *Resources, Conservation and Recycling*, Vol. 43 No. 4, pp. 419-432.

Einsiedel, E. and Eastlick, D. (2000), "Consensus conferences as deliberative democracy: a communications perspective", *Science Communication*, Vol. 21 No. 4, pp. 323-343.

Ely, A., Van Zwanenberg, P. and Stirling, A. (2014), "Broadening out and opening up technology assessment: approaches to enhance international development, co-ordination and democratisation", *Research Policy*, Vol. 43 No. 3, pp. 505-518.

Esch, M.E. (1972), "Honeywell's pattern: planning assistance through technical evaluation of relevance numbers", in Bright, J. and Schoeman, M.E.F. (Eds), *A Guide to Practical Technological Forecasting*, Prentice- Hall, Englewood Cliffs, NJ.

Evans, R., Brereton, D. and Joy, J. (2007), "Risk assessment as a tool to explore sustainable development issues: lessons from the Australian coal industry", *International Journal of Risk Assessment and Management*, Vol. 7 No. 5, pp. 607-619.

Farquhar, J. (1970), *A Preliminary Inquiry into the Software Estimation Process*, Rand, Santa Monica, CA.

Finsterbusch, K. (1980), *Understanding Social Impacts*, Sage Publications, Beverly Hills, CA.

Firat, A.K., Woon, W.L. and Madnick, S. (2008), "Technological forecasting: a review", Composite Information Systems Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

Fisher, J. and Pry, R. (1971), "A simple substitution model of technological change", *Technological Forecasting and Social Change*, Vol. 3, pp. 75-88.

Fleischer, T. and Grunwald, A. (2008), "Making nanotechnology developments sustainable: a role for technology assessment?", *Journal of Cleaner Production*, Vol. 16 Nos 8-9, pp. 889-898.

Flynn, R. and Bellaby, P. (2007), *Risk and the Public Acceptance of New Technologies*, Palgrave Macmillan, Basingstoke.

George, A.L. and Bennett, A. (2005), *Case Studies and Theory Development in the Social Sciences*, MIT Press, Cambridge, MA.

Gerring, J. (2004), "What is a case study and what is it good for?", *American Political Science Review*, Vol. 98 No. 2, pp. 341-354.

Gordon, T. and Helmer-Hirschberg, O. (1964), "Report on a long-range forecasting study", The RAND Corporation, Santa Monica, CA.

GRI (2006), *Sustainability Reporting Guidelines*, Global Reporting Initiative, Amsterdam.

Guimaraes, T. and Langley, K. (1994), "Developing innovation benchmarks: an empirical study", *Benchmarking for Quality Management & Technology*, Vol. 1 No. 3, pp. 3-20.

Gustafson, D.H., Shukla, R.K., Delbecq, A. and Walster, G.W. (1973), "A comparative study of differences in subjective likelihood estimates made by individuals, interacting groups, Delphi groups, and nominal groups", *Organizational Behaviour and Human Performance*, Vol. 9 No. 2, pp. 280-291.

Haque, E., Doja, S. and Haleem, A. (2013), "Technology forecasting: Indian perspective", *International Journal of Sustainable Development and Green Economic*, Vol. 2 No. 1, pp. 6-10.

Hill, K.Q. and Fowles, J. (1975), "The methodological worth of the Delphi forecasting technique", *Technological Forecasting and Social Change*, Vol. 7 No. 2, pp. 179-192.

Hoffmann, V.H., McRae, G.J. and Hungerbühler, K. (2004), "Methodology for early-stage technology assessment and decision making under uncertainty: application to the selection of chemical processes", *Industrial & Engineering Chemistry Research*, Vol. 43 No. 15, pp. 4337-4349.

Howitt, R. (1989), "Social impact assessment and resource development: issues from the Australian experience", *The Australian Geographer*, Vol. 20 No. 2, pp. 153-166.

Huang, L., Zhang, Y., Guo, Y., Zhu, D. and Porter, A.L. (2014), "Four-dimensional science and technology planning: a new approach based on bibliometrics and technology road mapping", *Technological Forecasting and Social Change*, Vol. 81, pp. 39-48.

Jain, R.K., Urban, L.V., Stacey, G.S. and Balbach, H.E. (1993), *Environmental Assessment*, McGraw-Hill Inc., New York, NY.

Jantsch, E. (1967), *Technological Forecasting in Perspective*, OECD, Paris.

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

Jeswani, H.K., Azapagic, A., Schepelmann, P. and Ritthoff, M. (2010), "Options for broadening and deepening the LCA approaches", *Journal of Cleaner Production*, Vol. 18 No. 2, pp. 120-127.

Jischa, M.F. (1998), "Sustainable development and technology assessment", *Chemical Engineering & Technology*, Vol. 21 No. 8, pp. 629-636.

Jolly, D.R. (2008), "Chinese vs. European views regarding technology assessment: convergent or divergent?", *Technovation*, Vol. 28 No. 12, pp. 818-830.

Jun, S. (2011), "A forecasting model for technological trend using unsupervised learning", in Kim, T. *et al.* (Eds), *Database Theory and Application, Bio-Science and Bio-Technology*, Springer, Berlin and Heidelberg, pp. 51-60.

Kerr, N.L. and Tindale, R.S. (2011), "Group-based forecasting? A social psychological analysis", *International Journal of Forecasting*, Vol. 27 No. 1, pp. 14-40.

Khurana, S., Mannan, B. and Haleem, A. (2014), "Integrating innovation with sustainability: a study of practices/status for Indian manufacturing industries (SMEs)", *Proceedings of AGBA's 11th World Congress Conference, Indian Institute of Technology, New Delhi, November 21-23.*

Kivikunnas, S. (1998), "Overview of process trend analysis methods and applications", ERUDIT Workshop on Applications in Pulp and Paper Industry, Sauli Kivikunnas University, Oulu.

Kleinman, D.L., Powell, M., Grice, J., Adrian, J. and Lobes, C. (2007), "A toolkit for democratizing science and technology policy: the practical mechanics of organizing a consensus conference", *Bulletin of Science, Technology & Society*, Vol. 27 No. 2, pp. 154-169.

Kumar, S., Luthra, S. and Haleem, A. (2015), "Benchmarking supply chains by analyzing technology transfer critical barriers using AHP approach", *Benchmarking: An International Journal*, Vol. 22 No. 4, pp. 538-558.

Levary, R.R. and Han, D. (1995), "Choosing a technological forecasting method", *Industrial Management-Chicago then Atlanta*, Vol. 37, pp. 14-24.

Madnick, S. and Woon, W.L. (2009), "Technology forecasting using data mining and semantics", MIT/MIST Collaborative Research, Cambridge, MA.

Martino, J. (2003), "A review of selected recent advances in technological forecasting", *Technological Forecasting and Social Change*, Vol. 70 No. 8, pp. 719-733.

Martino, J.P. (1970), "The precision of Delphi estimates", *Technological Forecasting*, Vol. 1 No. 3, pp. 293-299.

Mishra, S., Deshmukh, S.G. and Vrat, P. (2002), "Matching of technological forecasting technique to a technology", *Technological Forecasting and Social Change*, Vol. 69 No. 1, pp. 1-27.

Moriarty, J.P. and Smallman, C. (2009), "En route to a theory of benchmarking", *Benchmarking: An International Journal*, Vol. 16 No. 4, pp. 484-503.

National Resources Committee (1937), "Technological trends and national policy, including the social implications of new inventions", report of the National Resources Committee on the Social Implications of New Inventions, Washington, DC.

Nikolopoulos, K., Litsa, A., Petropoulos, F., Bougioukos, V. and Khammash, M. (2015), "Relative performance of methods for forecasting special events", *Journal of Business Research*, Vol. 68 No. 8, pp. 1785-1791.

Norgate, T.E., Jahanshahi, S. and Rankin, W.J. (2007), "Assessing the environmental impact of metal production processes", *Journal of Cleaner Production*, Vol. 15 No. 8, pp. 838-848.

Parente, F.J., Anderson, J.K., Myers, P. and O'brien, T. (1984), "An examination of factors contributing to Delphi accuracy", *Journal of Forecasting*, Vol. 3 No. 2, pp. 173-182.

Pfeiffer, J. (1968), *New Look at Education: System Analysis in Our Schools and Colleges*, Odyssey, New York, NY.

Phillips, J., Heidrick, T. and Potter, I. (2007), "Technology futures analysis methodologies for sustainable energy technologies", *International Journal Innovation Technology Management*, Vol. 4 No. 2, pp. 171-190.

Porter, A.L. (2010), "Technology foresight: types and methods", *International Journal of Foresight and Innovation Policy*, Vol. 6 Nos 1-3, pp. 36-45.

Raven, R.P., Jolivet, E., Mourik, R.M. and Feenstra, Y.C. (2009), "ESTEEM: managing societal acceptance in new energy projects: a toolbox method for project managers", *Technological Forecasting and Social Change*, Vol. 76 No. 7, pp. 963-977.

Reger, G. (2001), "Technology foresight in companies: from an indicator to a network and process perspective", *Technology Analysis & Strategic Management*, Vol. 13 No. 4, pp. 533-553.

Roberts, E. (1969), "Exploratory and normative technological forecasting: a critical appraisal", *Technological Forecasting*, Vol. 1 No. 2, pp. 113-127.

Roussel, P.A., Saad, K.N. and Erickson, T.J. (1991), *Third Generation R&D: Managing the Link to Corporate Strategy*, Arthur D. Little, Boston, MA.

Rowe, G. and Wright, G. (1999), "The Delphi technique as a forecasting tool: issues and analysis", *International Journal of Forecasting*, Vol. 15 No. 4, pp. 353-375.

Schnaars, S.P. (1989), *Megamistakes: Forecasting and the Myth of Rapid Technological Change*, Free Press, New York and London, p. 202.

Sharif, A.M. (2002), "Benchmarking performance management systems", *Benchmarking: An International Journal*, Vol. 9 No. 1, pp. 62-85.

Slocum, M.S. and Lundberg, C.O. (2001), "Technology forecasting: from emotional to empirical", *Creativity and Innovation Management*, Vol. 10 No. 2, pp. 139-152.

Slootweg, R., Vanclay, F. and van Schooten, M. (2001), "Function evaluation as a framework for the integration of social and environmental impact assessment", *Impact Assessment and Project Appraisal*, Vol. 19 No. 1, pp. 19-28.

Stewart, M. (1999), *Environmental Life Cycle Considerations for Design Related Decision Making in Minerals Processing*, University of Cape Town, Cape Town.

Stewart, M. (2001), "The application of life cycle assessment to mining, minerals and metals", report of the MMDS Workshop on Life Cycle Assessment, MMSD Project of IIED, New York, NY, 9-10 August.

Stewart, M. and Petrie, J. (2006), "A process systems approach to life cycle inventories for minerals: South African and Australian case studies", *Journal of Cleaner Production*, Vol. 14 No. 12, pp. 1042-1056.

Tasopoulou, K. and Tsiotras, G. (2017), "Benchmarking towards excellence in higher education", *Benchmarking: An International Journal*, Vol. 24 No. 3, pp. 617-634.

Technology futures analysis (2004), "Toward integration of the field and new methods", *Technological Forecasting and Social Change*, Vol. 71 No. 3, pp. 287-303.

Tran, T.A. and Daim, T. (2008), "A taxonomic review of methods and tools applied in technology assessment", *Technological Forecasting and Social Change*, Vol. 75 No. 9, pp. 1396-1405.

Twiss, B. (1984), "Technological forecasting for decision making", *Futures*, Vol. 16 No. 2, pp. 197-199.

Un, S. and Price, N. (2007), "Bridging the gap between technological possibilities and people: involving people in the early phases of technology development", *Technological Forecasting and Social Change*, Vol. 74 No. 9, pp. 1758-1772.

Van den Hende, E.A., Schoormans, J.P., Morel, K.P., Lashina, T., van Loenen, E. and de Boevere, E.I. (2007), "Using early concept narratives to collect valid customer input about breakthrough technologies: the effect of application visualization on transportation", *Technological Forecasting and Social Change*, Vol. 74 No. 9, pp. 1773-1787.

Wood, F.B. (1997), "Lessons in technology assessment: methodology and management at OTA", *Technological Forecasting and Social Change*, Vol. 54 No. 2, pp. 145-162.

Woudenberg, F. (1991), "An evaluation of Delphi", *Technological Forecasting and Social Change*, Vol. 40 No. 2, pp. 131-150.

Zhu, D. and Porter, A.L. (2002), "Automated extraction and visualization of information for technological intelligence and forecasting", *Technological Forecasting and Social Change*, Vol. 69 No. 5, pp. 495-506.

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

Further reading

Ayres, R. (1989), "The future of technological forecasting", *Technological Forecasting and Social Change*, Vol. 36 Nos 1-2, pp. 49-60.

Finsterbusch, K. and Motz, A. (1980), *Social Research for Policy Decisions*, Wadsworth, Belmont, CA.

Finsterbusch, K., Llewellyn, L. and Wolf, C. (1983), *Social Impact Assessment Methods*, Sage Publications, Beverly Hills, CA.

Goldfarb, R., Stekler, H. and David, J. (2005), "Methodological issues in forecasting: insights from the egregious business forecast errors of late 1930", *Journal of Economic Methodology*, Vol. 12 No. 4, pp. 517-542.

Haleem, A., Luthra, S., Mannan, B., Khurana, S., Kumar, S. and Ahmad, S. (2016), "Critical factors for the successful usage of fly ash in roads & bridges and embankments: analyzing Indian perspective", *Resources Policy*, Vol. 49, pp. 334-348.

Koivisto, R., Wessberg, N., Eerola, A., Ahlqvist, T., Kivisaari, S., Myllyoja, J. and Halonen, M. (2009), "Integrating future-oriented technology analysis and risk assessment methodologies", *Technological Forecasting and Social Change*, Vol. 76 No. 9, pp. 1163-1176.

Krishnamurthy, P., "Methodological premises of social forecasting in the context of business organizations", *SSRN Electronic Journal*, available at: http://dx.doi.org/10.2139/ssrn.1632513

Linstoni, H. (1969), "Technological forecasting and long-range planning", *Technological Forecasting*, Vol. 1 No. 2, pp. 227-228.

Mannan, B., Khurana, S. and Haleem, A. (2016), "Modeling of critical factors for integrating sustainability with innovation for Indian small- and medium-scale manufacturing enterprises: an ISM and MICMAC approach", *Cogent Business & Management*, Vol. 3 No. 1, pp. 1-5.

Medford, R. (1973), "A guide to practical technological forecasting", in Bright, J.R. and Schoeman, M.E.F. (Eds), *Futures*, Vol. 5, The University of Michigan, Michigan, pp. 590-592.

Slovic, P. (1999), "Trust, emotion, sex, politics, and science: surveying the risk-assessment battlefield", *Risk Analysis*, Vol. 19 No. 4, pp. 689-701.

Tallis, B., Azapagic, A., Howard, A., Parfitt, A., Duff, C. and Hadfield, C. (2003), *The Sustainability Metrics*, Institute of Chemical Engineering, Rugby.

Twiss, B. (1976), "Economic perspectives of technological progress", *Futures*, Vol. 8 No. 1, pp. 52-63.

van Andel, T. and Aston, M. (1987), "Interpreting the landscape-landscape archaeology in local studies", *American Journal of Archaeology*, Vol. 91 No. 3, pp. 489-490.

About the authors

Dr Abid Haleem is Professor and Ex-Head of Mechanical Engineering Department, Faculty of Engineering and Technology, and is also Honorary Director Internal Quality Assurance Cell (IQAC) Jamia Millia Islamia (A Central University by an Act of Parliament), New Delhi, India. Professor Haleem received the PhD Degree from IIT (Delhi) in the area of "Strategic Management". He completed his Graduation and Post-Graduation Degrees in "Mechanical Engineering" and "Industrial Engineering", respectively. He was awarded Gold medal at post-graduation level. Professor Haleem has more than 100 research papers to his credit, published in international and national journals. He has authored a book titled *Innovation, Flexibility and Technology Transfer*, published by Tata McGraw Hill, India. He is also Regional Editor Asia Pacific Region, for *Global Journal of Flexible Systems Management*. He has visited academic institutions and Industries at the USA, the UK and Australia. He has extensive experience in coordinating different types of academic programs in the field of management and technology. He has experience in industrial automation products of Rockwell Automation, Milwaukee, WI, USA. He has also consulting experience with some international as well as national organisations. Some of them are Department for International Development, Educational Consultants, Government of Orissa, Government of Aizwal, ESCORTS Group, AICTE, etc. He carries rich industrial experience and has completed AICTE projects worth one million Indian rupees in the field of industrial engineering. He has also visited institutions at Australia, UK and USA. He was on the Board of Telecommunications Consultants India Limited as Independent Director from 2008 to 2011. He is also involved in various institutional activities (engineering and management institutions and universities) related to policy planning, administration, accreditation, curriculum design, admission, evaluation and examination process, etc. He is actively involved in teaching as well as research work at graduate, postgraduate and PhD levels in the areas of Technology Management, Innovation, industrial engineering, green practices, supply chain management, business process re-engineering, information resource management, e-governance, technology management and allied areas. He has produced eight PhD and seven students are enroled with him for their research work under his guidance.

Bisma Mannan is Research Scholar in Jamia Millia Islamia. She has published a book with LAP Publication, Germany. Her papers were published in three national conferences and six International conferences. One of her paper published in *IJCA*, USA. Her research interests include innovation, green practices, knowledge management, technology management, robotics applications area among others.

Dr Sunil Luthra is working as Assistant Professor, State Institute of Engineering and Technology (formerly known as Government Engineering College), Nilokheri, Haryana, India. He has been associated with teaching for the last 15 years. He has contributed over 100 research papers in international referred and national journals, and conferences at the international and national level. His scholarly work has also been acknowledged in several International journals of repute such as the *JCP*, *IJPE*, *PPC*, *IJPR*, *RSER*, *RCR*, *EGY*, *JRPO* and many more, and conference of repute, such as SOM-14, NITIE – POMS, AGBA, GLOGIFT 14 and GLOGIFT 15, etc. His research is in the spotlight. His works got more than 1,500 citations (h-index = 20). His RG score is higher than 87.5 per cent of Research Gate members. His specific areas of interest are operation management, green supply chain management, sustainable supply chain management, sustainable consumption and production, reverse logistics, renewable/sustainable energy technologies and business sustainability, etc. Dr Sunil Luthra is the corresponding author and can be contacted at: sunilluthra1977@gmail.com

Dr Sanjay Kumar is working as Professor in Mechanical Engineering Department, I.I.T.M, Sonepat, India. He received the PhD Degree in Mechanical Engineering from Jamia Millia Islamia University Delhi, India. He did Masters in Mechanical Engineering with specialisation in Production Engineering from Delhi Technological University, Delhi, India, and Bachelor Degree in Mechanical Engineering from National Institute of Technology, Kurushetra, India. He has been associated with industry and teaching in the various fields of Mechanical Engineering and Operation Management for over 15 years. He has contributed over 50 research papers in international referred and national journals, and conferences at international and national level. His specific areas of interest are industrial engineering, operations management, supply chain management, technology transfer, green supply chain management, etc.

Sonal Khurana is Research Scholar in Jamia Millia Islamia. She received the BTech Degree in Mechanical and Automation Engineering from GGSIPU and the Mtech Degree in Industrial and Production Engineering from Jamia Millia Islamia. She was awarded Gold medal at her MTech level. She has been teaching in a college affiliated to GGSIPU for the past two years. Her papers were published in three national conferences and six international conferences. One of her papers got published in *IJCA*, USA. Her research interests include green practices, sustainable-oriented innovation, sustainable manufacturing, technology management, robotics applications area among others.

Downloaded by Delhi Technological University At 22:00 05 January 2019 (PT)

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**

Transportation Research Part D 67 (2019) 541–552

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

The effect of odd-even driving scheme on $PM_{2.5}$ and $PM_{1.0}$ emission

Rajeev Kumar Mishra[a,*], Abhinav Pandey[a], Govind Pandey[b], Amrit Kumar[a]

[a] Department of Environmental Engineering, Delhi Technological University, Delhi 110 042, India
[b] Civil Engineering Department, Madan Mohan Malaviya University of Technology, Gorakhpur 273 010, India

ARTICLE INFO

Keywords:
Low Emission Zones
Odd-even scheme
Urban air pollution
Vehicular emission

ABSTRACT

According to a global survey of 1600 cities by WHO in 2015, Delhi was found to be the most polluted city in the world, highlighting immediate need to bring in appropriate and effective measures to ensure a healthy air quality throughout the city. In a bid to do so, the odd-even driving scheme was implemented a few years ago in Delhi city on a trial basis requiring the vehicles bearing odd and even registration numbers to run on alternative days with effect from 1st January 2016. Applied for the very first time in India and lasting on a pilot basis for 15 days and also referred to as Phase 1 of implementation, the scheme ended on 15th January 2016 with primary goal of assessing the extent to which such a scheme could serve as an effective measure to bring down and control the excessively high ambient air pollution levels. Other associated benefits, which the study considered, were decongestion of city roads, and promoting use of public transport system and carpooling, besides other alternative modes of transport. The paper presents a comparative analysis of particulate matter concentrations recorded before and during the implementation of odd-even scheme of Phase 1 for three key traffic corridors of the city. The study reports an average reduction of $PM_{2.5}$ of 5.73% and that of $PM_{1.0}$ of 4.70% in the ambient air across all the corridors during the days of implementation of scheme, which depicts a small but positive impact of the measure adopted towards abatement of urban air pollution in the megacity of Delhi.

Capsule: The study finds that the odd-even driving scheme can help to reduce particulate matter levels compared to non-scheme days at traffic corridors.

1. Introduction

The history of odd (number)-even (number) license plate based driving restriction dates back to 1979, when it was instituted in the United States as a part of responding to the second gasoline crisis, when conditions in Iraq and Iran contributed to crude oil price increases globally, and the supply of gasoline was made available on alternate days. Though this was not done with the aim of reducing vehicular emission in the US, however, since then, various capitals and major cities worldwide have adopted an odd-even or restricted driving rule in one form or the other (often targeted at peak period commuter traffic) with twin goals to reduce the urban traffic congestion and the associated automobile exhaust emissions. Such driving restrictive measures were implemented in cities like Paris, Mexico, Bogota and Beijing, with different end-results following removal of the restrictions. It should also be noted that studies have found that over 60–70% of air pollution in developing countries, particularly, faster-than-ever growing megacities, is attributed to automobile emissions caused by a large fleet of older vehicles exhibiting higher age and mileage, poor inspection/maintenance history, inadequate road infrastructure and low fuel quality (Pandey et al., 2016a,b; Wang et al., 2010; Singh et al., 2007; Badami, 2005; Auto Fuel Policy, 2002).

India has experienced a dramatic growth in vehicle production, ownership, and usage. While 8.46 million vehicles were produced in 2004–2005, India recorded a production of 23.36 million vehicles in the year 2014–2015, thereby, registering a decennial growth rate of about 63.76%. It is found that passenger cars, three-wheelers and two-wheelers have witnessed high growth rates of 62.43%, 60.54% and 64.70% respectively during that decade, while total commercial vehicles recorded a 49.26% increase over the same decade in the country. Recent figures show that the number of vehicles has ballooned to more than 10 million in Delhi in 2017, with two-wheelers constituting the highest share at 64% followed by passenger vehicles including cars and jeeps at about 30% (Government of India, 2017).

Tremendous growth of the industrial sector has also been contributing to the deteriorating ambient air quality in Delhi. According to the 6th Economic Census 2013, there were approximately 0.89 million industrial units in operation in Delhi, registering an increase of 18.35% compared to the number of such units highlighted in the 5th Economic Census conducted in 2005 (ESD, 2015). Although, air pollution monitoring and control devices were found installed in most of the industrial units, it is of serious concern that many such devices and equipment are generally non-functional either due to poor handling or lack of adequate maintenance. It is also evident that not much attention is being paid to these control devices, in terms of periodic servicing or required repairs. Inappropriate height of installation of stacks in the industry premises was also found to be one of the key factors contributing to air pollutants escaping into ambient air adversely affecting atmospheric health in the vicinity. However, owing to the stringency of the emission controls that have forced the industrial establishment to comply with the same or to relocate themselves to the city outskirts, industrial sector's contribution to air pollution has reduced to 52%. On the other hand, the contribution of the automobile sector to air pollution increased to 36% in the year 2013–14 (Sharma and Dixit, 2016), primarily due to a very high increase in vehicle numbers in the city in the past several years.

Increasing traffic congestion on the city roads is another significant factor attributing to urban air pollution. As is evident, cars and jeeps are not only responsible for a major proportion of Delhi's vehicular population, but also for congestion on the roads. These vehicles need more parking space than two-wheelers, and roadside parking reduces available lane width of the road leading to traffic congestion and hence ground level ambient air concentration build-up. This build-up of automobile emission is further worsening in the megacity of Delhi, where the situation of ambient air quality has already been alarming for a number of years (Pandey et al., 2016a,b). In a recent study based on the data of SO_2, NO_2, and PM for 2009–2015, the exceedance factors (EF) were calculated and it was found that pollution levels for SO_2 corresponded to low category and the same for NO_2 remained in the high category. However, the pollution levels have been found in excess of critical level for PM in all years (CPCB-ENVIS, 2016). This scenario has prompted Government of Delhi to introduce the driving restriction scheme on trial basis.

It is interesting to note that the phenomenon of Urban Air Quality Management (UAQM), which has recently become a major thrust area for researchers worldwide, highlights various strategies to abate the ever-increasing vehicular emission in developing and developed countries. Strategies include driving restrictions of various kinds, introduction of LEZs (Low Emission Zones), ban on entry of old and primarily diesel-driven heavy goods vehicles in designated air quality control regions, ban on street parking, use of pricing schemes, and so on. Another key strategy being strongly backed by the Government of India is implementation of stringent vehicular emission norms, particularly for in-use vehicles. Currently, the country has BS (Bharat Stage) IV emission norms which are proposed to be replaced by BS VI norms skipping the 'V' version with effect from April 1, 2020. This is not only to keep abreast with otherwise much stringent EURO norms, but also, to quickly shift to availability of BS VI quality (a relatively cleaner gasoline that has only 10 ppm of Sulphur compared to 50 ppm in existing BS IV fuel). The odd-even driving scheme as proposed in case of Delhi city is one such abatement measure and hence an analysis of its potential effectiveness in reducing automobile emissions is worthy of investigation.

2. Literature survey

Driving restrictions, in a bid to either curb the dangerously high levels of ambient air pollution or to decongest the urban traffic, have been implemented in several cities and metropolitan areas all across the world. However, limited by weaknesses of the prevalent research methodology and/or data gaps, only a few studies have quantified the effects of such restrictions on traffic and associated air quality improvements, if any. Data measuring traffic condition was introduced for the first time in similar studies to help explain the mechanism of effect of driving restrictions on traffic and air quality in Beijing city. It was reported that though more stringent driving limitations had a positive impact on urban traffic speed, a relatively low impact was observed on concentration of inhalable particulates, which further implied that a policy change might not achieve positive effects for both traffic as well as environment. (Cong et al., 2014).

Yunxia et al. (2015), in their study on exploration of driving restriction policy acceptance in some large Chinese cities, reported that it was up to policymakers to ultimately not only improve public transport, but also, enhance commuters' awareness of outcomes, and promote car owners to change their driving behavior. Otherwise, a driving restriction policy may not be enough to motivate commuters to opt for public transport. Based on daily data from various monitoring stations, air pollution was reported to have been reduced by 21% during 'weekly-once' limitation in Beijing city (Viard and Shihe, 2015).

Davis (2008) assessed the effects of driving limitations on air quality using high-frequency measurements obtained from monitoring stations in Mexico City. The study reported no evidence that the restrictions had improved air quality across pollutants and specifications. The impact of low emission zone and bans on heavy-duty vehicles was investigated in Munich (Germany), using a semi-parametric regression model for modeling PM_{10} levels. The study reported a higher reduction in PM_{10} concentration at a traffic monitoring site (19.6% in summer, and 6.8% in winter) after the introduction of measures, and smaller reductions in urban

* Corresponding author.
E-mail address: rajeevkumarmishra@dtu.ac.in (R.K. Mishra).

https://doi.org/10.1016/j.trd.2019.01.005

1361-9209/ © 2019 Elsevier Ltd. All rights reserved.

background (5.7% in summer, and 3.2% in winter) (Fensterer et al., 2014).

Gramsch et al. (2013) evaluated the impact of the Transantiago system on the black carbon (BC) pollution along four roads, which were directly affected by the alterations to the transport system. The alteration was found to be related to a decrease in the total number of vehicles or the number of other diesel vehicles in the street rather than a decrease in the number of buses alone. This highlighted that reduction in total number of vehicles and not in buses only, can contribute to improved ambient air quality.

Based on a three day campaign, Invernizzi et al. (2011) found that the restricted area in Milan (referred to as Ecopass) had considerably lower black carbon concentrations than regular traffic areas, but the concentrations of PM_{10}, $PM_{2.5}$ and $PM_{1.0}$ did not show any significant difference. Also, the ratio of black carbon to PM_{10} was reported to be decreased by 47% and 62% by the authors in the Ecopass zone and in the pedestrian zone, respectively, as compared to the no-restriction zone.

Boogaard et al. (2012) investigated effects of LEZs' implementation on air quality in five Dutch cities and found that the LEZ policies directed at older heavy-duty trucks did not bring about the desired result of substantial decrease in the concentration of related air pollutants due to these vehicles forming only a small part of overall fleet.

Cyrys et al. (2014) observed that PM_{10} concentrations decreased by 10% after the LEZ implementation in Colonge, Berlin and Munich. Titos et al. (2015) analyzed the effect of different driving restrictions on air quality in two similar cities and reported a significant reduction of 72% in black carbon (BC) concentration and 37% reduction in BC and 33% reduction in PM_{10} concentrations in cities of Granada (Spain) and Ljubljana (Slovenia) respectively during restriction days, attributing such reduction to re-organization of public transportation, renewal of the bus fleet and closing streets to private traffic.

A substantial reduction in traffic-related concentration of pollutants in the vicinity of a roadside monitoring station after the implementation of a LEZ in Amsterdam was reported by Panteliadis et al. (2014) with a relatively sharper decrease in elemental carbon (EC) values. Qadir et al. (2013) found a positive impact of implementation of LEZ in Munich, reporting a 60% decrease in contribution of traffic source after the implementation and a reduction in the average concentration of EC from traffic factor from 1.1 to 0.5 $\mu g/m^3$.

It is seen that in most of the studies, the quantification of PM emissions in the ambient air quality under various types of driving restrictions (mostly on LEZs) has been reported. However, no study has assessed the $PM_{2.5}$ and $PM_{1.0}$ concentrations and their pattern of varying ratios for different traffic corridors in the context of an odd-even day driving scheme so far in India.

It is worthwhile to mention here that particles of size between 1 and 2.5 μm show a substantial retention in the bronchioles and alveoli of human respiratory system. Even if particulate matter ranging from size 0.25 to 1 μm may show a decrease in retention, particles below 0.25 μm size often show an increase in retention because of Brownian motion. Also, $PM_{1.0}$ is considered to be a good indicator for vehicular emissions at roadside locations (Lee et al., 2006).

The air pollution-induced human health impacts have been reported by various researchers world-wide; in particular, those sourced from road transport are demonstrated to be not the only factor behind increased risk of respiratory and cardiovascular diseases including cancer (Pope and Dockery, 2006), but also overall mortality increase, lung cancer cases and various other respiratory ailments (Brauer et al., 2002). The World Health Organization (WHO), vide its Urban Ambient Air Pollution database - Update 2016, has reported that as many as 50% of the cities, especially those in south-east Asia, Eastern Mediterranean and Western Pacific regions have recorded an increasing trend of PM_{10} or $PM_{2.5}$ values over the last 5 year span. It also estimated around 7 million deaths (one in eight of total global deaths) as a result of air pollution exposure, which not only more than doubled compared to previous findings, but also confirmed air pollution to be world's largest single environmental health risk (WHO, 2005, 2014, 2016).

The situation is reported to be even more acute in many large urban areas of fast-growing economies, particularly, those in China and India, where concentrations are reported to be several times higher than corresponding WHO guidelines. Such higher levels are due to a combination of pollutants from a very wide variety of emission sources, mainly, industry, transportation, construction work (dust) and open-air waste burning (IEA, 2016).

With this in view, a study was undertaken in the megacity of Delhi involving data acquisition and analysis for assessing the relevance and effects of odd-even driving scheme on PM emissions at different traffic corridors. In the present work, in view of the limitations of time and resources, an attempt was made to quantify the effects of odd-even driving restriction scheme implemented in Phase 1 on three key traffic corridors of Delhi city on real-time particulate matter data ($PM_{2.5}$ and $PM_{1.0}$) collected one day prior to introduction of the scheme, followed by two days of continuous monitoring of 12 h during the span of the scheme covering all types of vehicles.

3. Materials and methods

3.1. Features of odd-even driving scheme in Delhi

Deluged by the outcomes of various pertinent national and international studies reporting the megacity of Delhi to be one of the most polluted cities in the world with air quality being indicated as 'bad or worse or severe' almost throughout the year, the odd-even driving scheme was implemented in India's capital city on an experimental basis. The scheme required vehicles with odd and even numbered license plates to run on alternate days with effect from 1st January 2016 to 15th January 2016 as a measure to bring down and control the excessively high ambient air pollution levels in the city. The restriction scheme defined the vehicles that can be driven on a particular date. On odd dates, only cars with license plates bearing odd number were allowed to ply on city roads and vice-versa. For instance, on dates ending with 0, 2, 4, 6 and 8, only those vehicles whose vehicle registration number ended with an even digit (such as DL12 CL 5672) were allowed to ply on the roads; and on odd dates, vehicles whose registration number ended with an odd number (for example, DL12 CL 5673) were permitted to be operated on city roads.

The first pilot or Phase 1 of implementation of this scheme began on January 1, 2016 and lasted for 15 days, before ending on January 15, 2016. The restrictions were in effect during this period from Monday to Saturday between 8 am and 8 pm in the megacity of Delhi. The scheme applied to the general population, with top level government bureaucrats and functionaries being exempted from the scheme. However, there were some exemptions for the general public as well; for instance, CNG (Compressed Natural Gas), LPG (Liquefied Petroleum Gas) vehicles, electric vehicles, hybrid vehicles, two-wheelers, women car drivers with only women and/or children, hospital vehicles, vehicles with physically challenged person, fire or medical emergency vehicles, law enforcement vehicles, and vehicles carrying top government officials were exempt from the scheme.

While violators of the odd-even scheme had to pay a fine of INR 2000 (Indian National Rupees; 65 INR = 1 USD (in 2015–2016) approximately. The amount of fine was kept high with a view to ensure better compliance. As a part of government initiatives, the road transport department ran close to 3000 extra buses (50% of those being diesel and remaining CNG-driven) on various roads of the capital city whereas the metro rail system increased the frequency of trains on all its routes (running 198 trains to make additional 365 trips) every day for ensuring a seamless and hassle-free experience for daily commuters. In addition to the above and with a view to ease the pressure on public transport facilities, the popular private taxi service providers in the city started offering car-pooling services for commuters during odd-even scheme days.

3.2. Study areas

Three major traffic corridors in Delhi, namely, Pitampura (Madhuban Chowk), Panchkuian Road and Najafgarh Road were selected. Real-time particulate matter data ($PM_{2.5}$ and $PM_{1.0}$) were collected along these three corridors prior to and during Phase 1 of implementation of odd-even driving scheme in the city. The corridors were chosen on the basis of minimum 24-hour average vehicle count being higher than 70,000, having a good mix of different vehicle types, and also considering the fact that these were main connecting corridors between the subsequent districts of the capital city. For example, Madhuban Chowk connects North-West District's traffic to other neighboring Outer Districts whereas Najafgarh Road connects major traffic from South-West District to South Delhi including major traffic from adjacent airport terminals. Furthermore, the corridors were also selected in such a way that they do not cover only the commercial and residential areas but take into account the contribution to PM from other land-use patterns, (e. g., industrial and/or construction sites).

3.3. Traffic characteristics

The data in respect of fleet characteristics for the referred corridors are presented in Table 1. It shows 24-hour average traffic count and vehicular composition collected by manual counting in the months of December 2015 and January, 2016.

3.4. Monitoring tools/instrumentation

The Haz-Dust EPAM-5000, a high sensitivity particulate monitor was used to monitor real-time $PM_{2.5}$ and $PM_{1.0}$ levels in the ambient air along the selected traffic corridors. This device is coupled with filtering techniques and real-time monitoring methods. It is accompanied by Dust Comm Pro software which helps in compiling and analyzing sampled data, and easily exporting the data in comma-delimited ASCII Text Files further usable in spreadsheet programs such as Microsoft Excel. The software enables its users to conduct a relatively detailed statistical analysis, create graphics and charts, and perform mathematical correction of particle characteristics (if required).

3.5. Monitoring schedule

A monitoring schedule was prepared covering three major traffic corridors of Delhi city for a 12 h period of continuous monitoring starting from 8:00 hrs to 20:00 hrs one day before and two days during Phase 1 of implementation of the odd-even driving scheme and spanned between 22nd December 2015 and 15th January 2016. The schedule of PM monitoring in terms of different corridors and dates is given in Table 2.

Table 1
Traffic characteristics of the selected transport corridors.

S. No.	Monitoring Station/Traffic Corridor	24-hrs. average count (December 2015 – January 2016)								
		Cars/Taxi	2-Ws	ARs	Buses	3-Ws (Goods)	HCVs	MCVs	LCVs	Total
1	Pitampura – Madhuban Chowk	75,397	9567	39,471	4840	7318	4967	1177	10,720	153,457
2	Panchkuian Road	29,255	37,354	8787	2335	1264	2202	4782	7648	93,627
3	Najafgarh Road	19,146	33,386	6488	4857	1975	1962	363	5479	73,657

Note: 2-Ws: Two-wheelers; 3-Ws: Three-wheelers; ARs: Auto-Rickshaws; H/M/LCVs: Heavy/Medium/Light Commercial Vehicles.

Table 2
Schedule of $PM_{2.5}$ and $PM_{1.0}$ monitoring in Delhi city.

S. No.	Monitoring Station/Traffic Corridor	PM monitoring schedule – Pre Odd-even scheme (before 1st Jan. '16)	PM monitoring schedule – During Odd-even scheme (1st – 15th Jan. '16)
		Date(s)	
1	Pitampura – Madhuban Chowk	22nd Dec. '15	12th and 13th Jan. '16
2	Panchkuian Road	23rd Dec. '15	8th and 11th Jan. '16
3	Najafgarh Road	30th Dec. '15	14th and 15th Jan. '16

4. Data analysis, results and discussion

The monitoring results of $PM_{2.5}$ and $PM_{1.0}$ at all the three monitoring stations/traffic corridors, namely, Pitampura (Madhuban Chowk), Panchkuian Road and Najafgarh Road are depicted in Figs. 1–3 and 4–6, and $PM_{1.0}$ to $PM_{2.5}$ ratios for the same are shown in Figs. 7–9.

4.1. $PM_{2.5}$

It is revealed from Figs. 1–3 that there is a reduction in $PM_{2.5}$ concentration during odd-even driving scheme days 1 and 2 at all the monitoring stations. It is also noteworthy that morning peak shows a decreasing trend in case of $PM_{2.5}$ concentrations till 14:00–15:00 h in the afternoon after which an increasing trend is observed. Further, a fall in concentrations is observed after 18:00–19:00 h.

These variations are attributable to the traffic conditions at monitoring stations with a general trend of increase or decrease in traffic counts with respect to time of the day.

The monitoring of temporal variation of $PM_{2.5}$ concentration reported a maximum observed value of 896 μg/m^3 at Panchkuian Road corridor, whereas the minimum was reported as 313 μg/m^3 at Najafgarh Road corridor before the odd-even scheme days. The same was found to be maximum as 873 μg/m^3 at Panchkuian Road and minimum as 217 μg/m^3 at Pitampura-Madhuban Chowk respectively during the scheme days. When compared to the corresponding 24-h NAAQS value of 60 μg/m^3, none of the three traffic

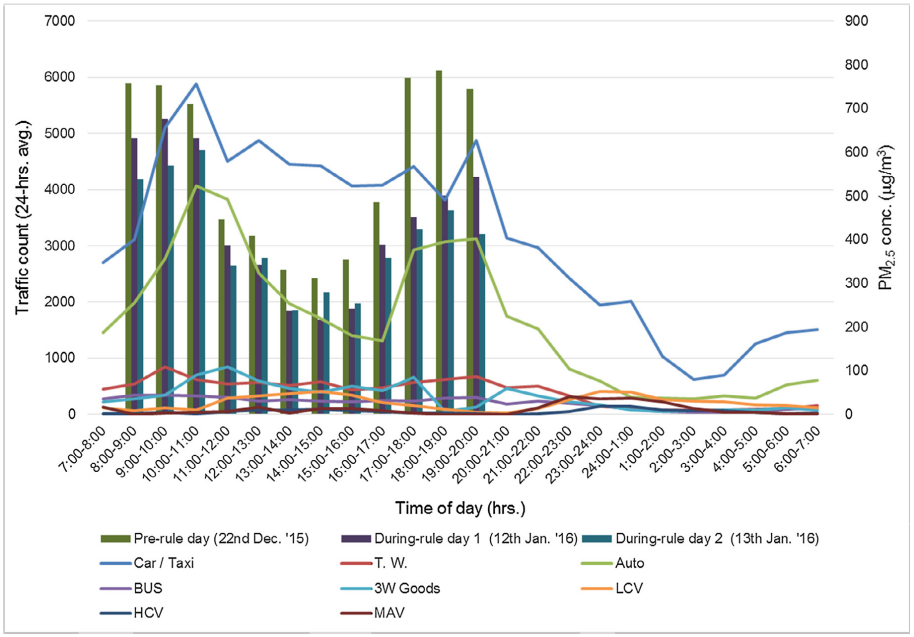


Fig. 1. Temporal variation of $PM_{2.5}$ concentration at Pitampura-Madhuban Chowk.

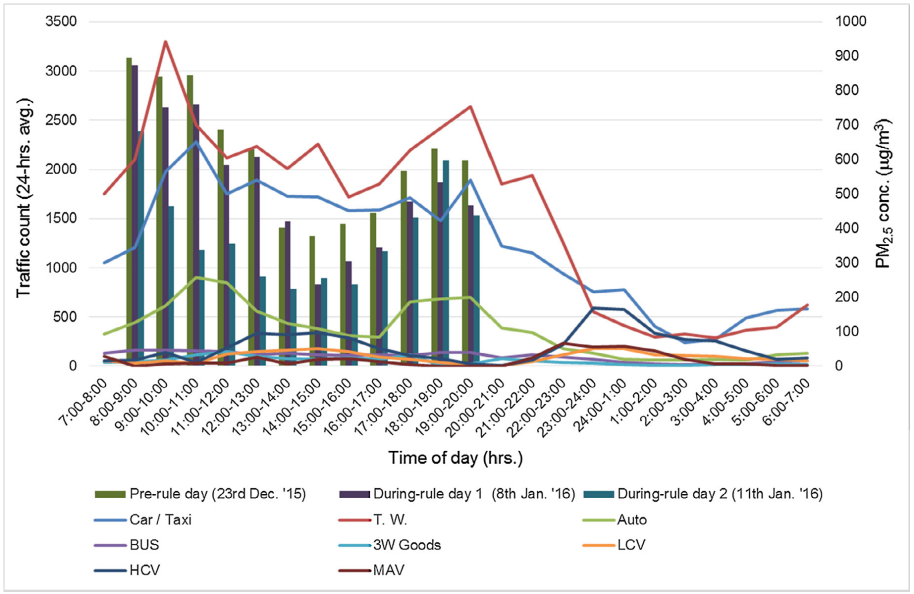


Fig. 2. Temporal variation of $PM_{2.5}$ concentration Panchkuian Road.

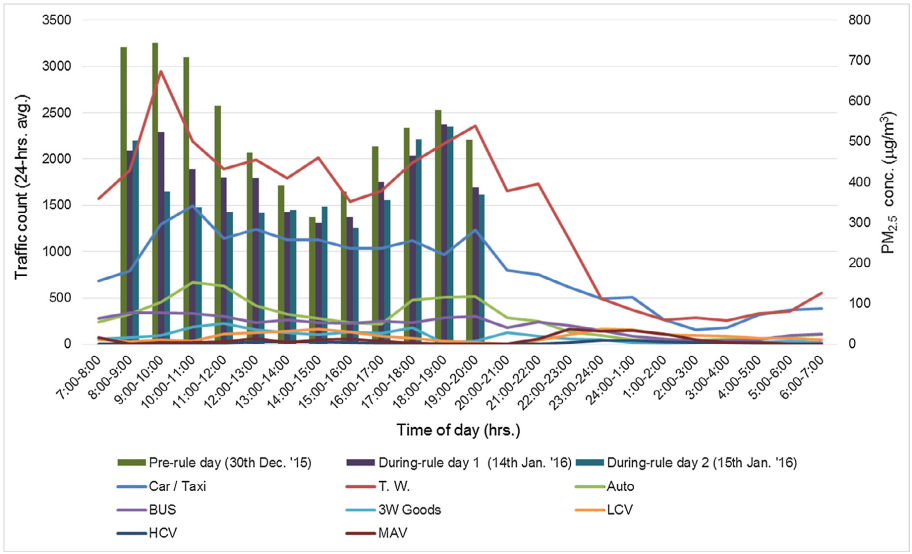


Fig. 3. Temporal variation of $PM_{2.5}$ concentration at Najafgarh Road.

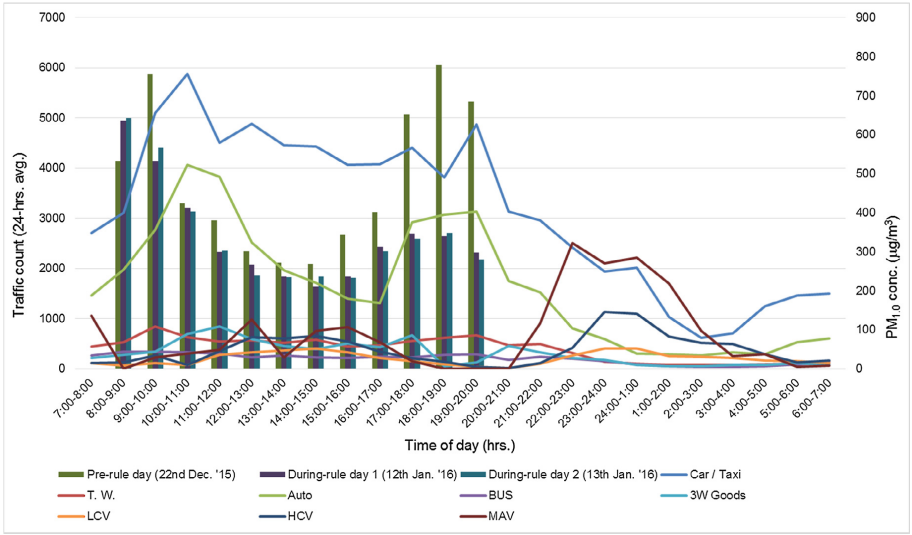


Fig. 4. Temporal variation of $PM_{1.0}$ concentration at Pitampura-Madhuban Chowk.

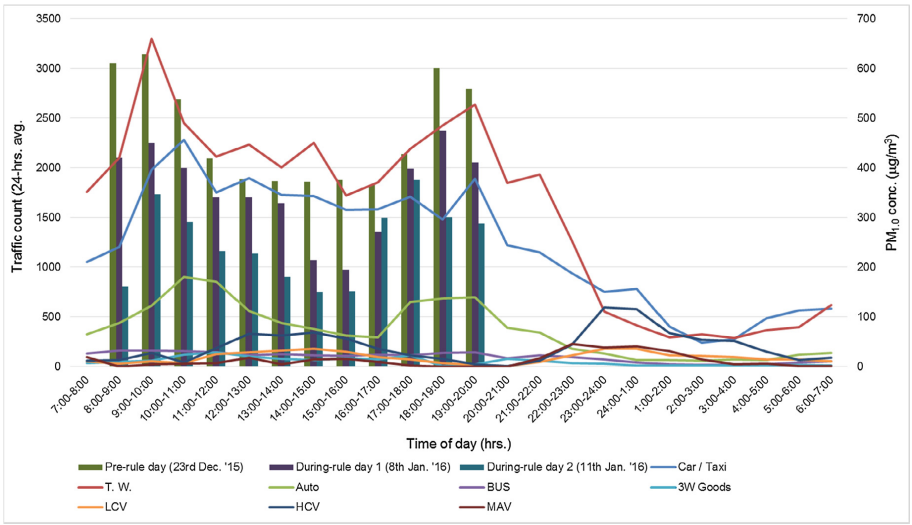


Fig. 5. Temporal variation of $PM_{1.0}$ concentration at Panchkuian Road.

corridors showed compliance at any given time, essentially remaining in non-compliance and depicting a value almost four times the NAAQS value even during the scheme day(s). Further, the maximum reduction of $PM_{2.5}$ concentration was reported as 8.96% at Panchkuian Road corridor and the minimum as 3.11% at Pitampura-Madhuban Chowk during the scheme days.

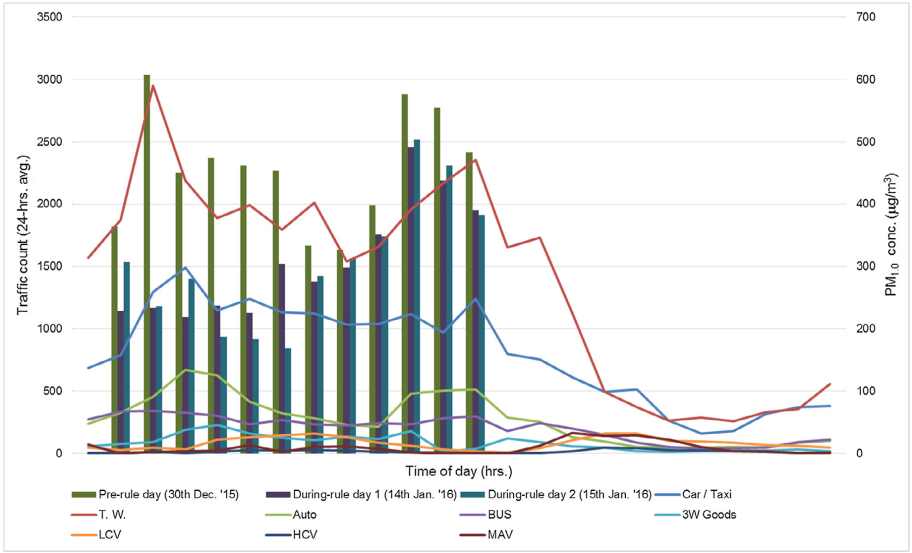


Fig. 6. Temporal variation of $PM_{1.0}$ concentration at Najafgarh Road.

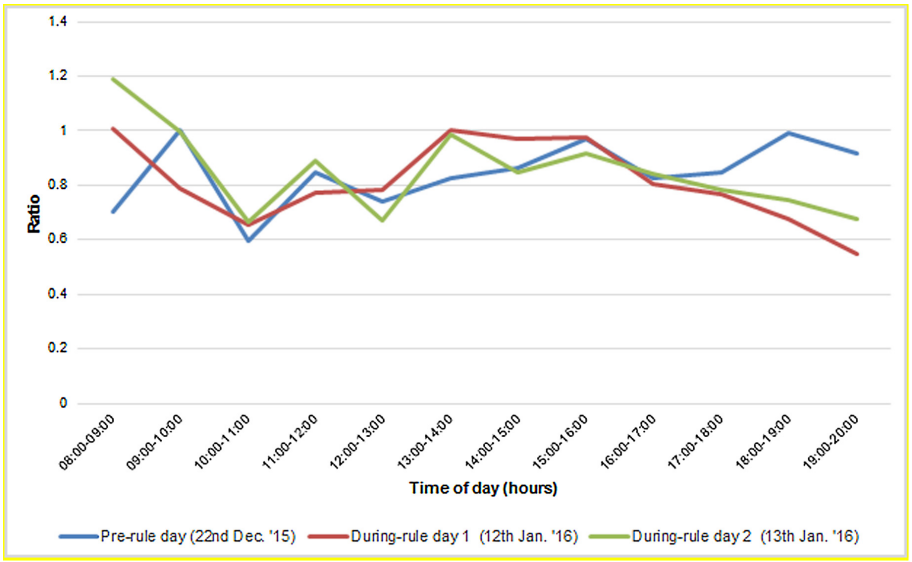


Fig. 7. $PM_{1.0}$:$PM_{2.5}$ ratio at Pitampura-Madhuban Chowk.

4.2. $PM_{1.0}$

A glance at Figs. 4–6 reflects that the odd-even scheme has a positive beneficial effect on the concentration of $PM_{1.0}$ implying that the observed value of this pollutant has decreased during the restriction days similar to the observation reported in case of $PM_{2.5}$.

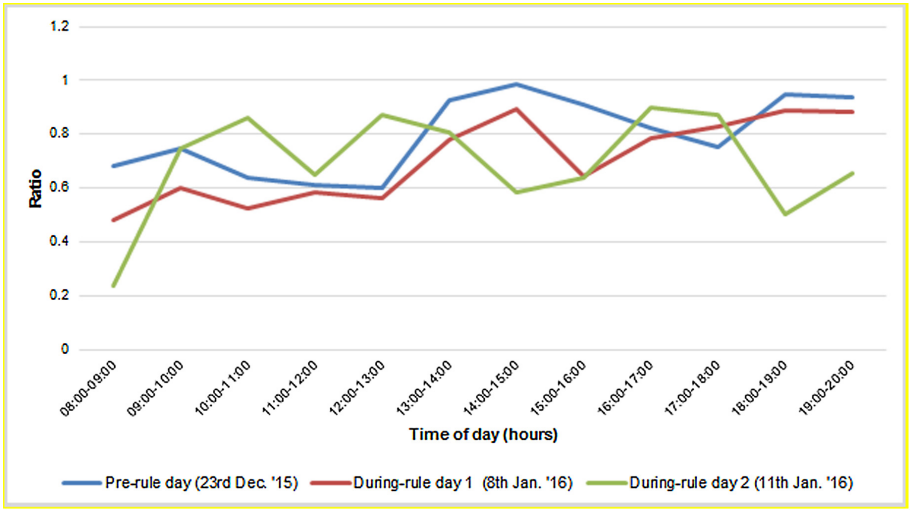


Fig. 8. $PM_{1.0}$:$PM_{2.5}$ ratio at Panchkuian Road.

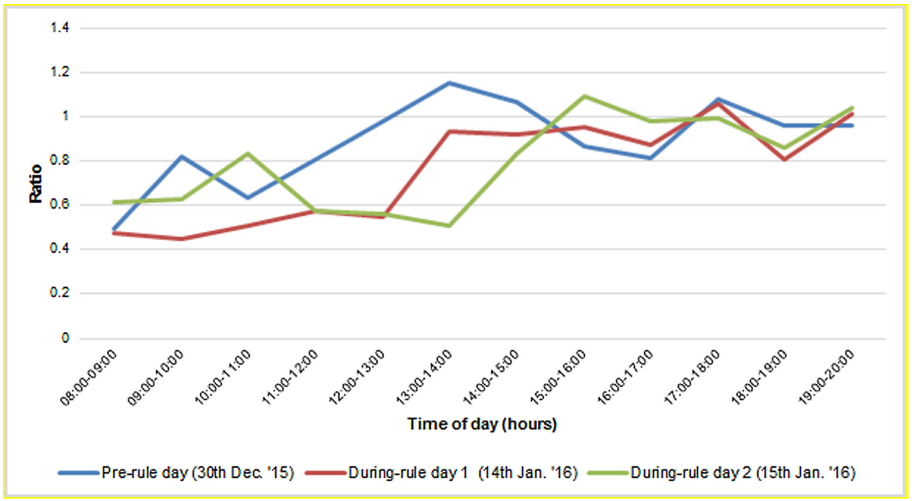


Fig. 9. $PM_{1.0}$:$PM_{2.5}$ ratio at Najafgarh Road.

While the maximum and minimum observed values of $PM_{1.0}$ were recorded as 756 μg/m^3 and 269 μg/m^3 both at Pitampura-Madhuban Chowk before the odd-even scheme day, the same was found to be maximum as 642 μg/m^3 at Pitampura-Madhuban Chowk and minimum as 150 μg/m^3 at Panchkuian Road respectively during the scheme days. A comparison of these values with 24-hour NAAQS value of 60 μg/m^3 for $PM_{2.5}$ (to be noted that a corresponding value for $PM_{1.0}$ is yet to be formulated in India), depicted no compliance at any given time even during the scheme day(s). Further, Pitampura-Madhuban Chowk showed both the maximum reduction of $PM_{2.5}$ concentration as 7.78%, whereas the minimum as 2.69% during the scheme days.

Table 3
Summary of $PM_{1.0}$ to $PM_{2.5}$ ratio at various traffic corridors.

S. No.	Monitoring Station/Traffic Corridor	$PM_{1.0}$:$PM_{2.5}$			
		Pre-scheme Day		During-scheme Days	
		Max.	Min.	Max.	Min.
1	Pitampura – Madhuban Chowk	1.00 (9:00am – 10:00am)	0.59 (10:00am – 11:00am)	1.19 (8:00am – 9:00am)	0.54 (7:00 pm – 8:00 pm)
2	Panchkuian Road	0.98 (2:00 pm – 3:00 pm)	0.59 (12:00 pm – 1:00 pm)	0.89 (4:00 pm – 5:00 pm)	0.43 (8:00am – 9:00am)
3	Najafgarh Road	1.15 (1:00 pm – 2:00 pm)	0.45 (8:00am – 9:00am)	1.09 (3:00 pm – 4:00 pm)	0.44 (8:00am – 9:00am)

4.3. *$PM_{1.0}$ To $PM_{2.5}$ ratios*

The $PM_{1.0}$ to $PM_{2.5}$ ratios were calculated in a bid to ascertain the proportion of $PM_{1.0}$ and $PM_{2.5}$ on all the three traffic corridors. The higher ratios are considered to be an indicator of presence of significant amount of vehicular exhaust. The average ratios of $PM_{1.0}$:$PM_{2.5}$ during scheme days were found to be 0.83, 0.7 and 0.78 whereas the same during pre-scheme days were reported as 0.84, 0.8 and 0.89 at monitoring stations/traffic corridors Pitampura-Madhuban Chowk, Panchkuian Road and Najafgarh Road respectively.

The obtained ratios of minimum 0.7 and maximum 0.83 during scheme days and 0.8 and 0.89 during pre-scheme days respectively being reasonably higher, indicate substantially higher presence of PM in the vicinity of traffic corridors. The average reductions in $PM_{1.0}$ to $PM_{2.5}$ ratio at monitoring stations/traffic corridors Pitampura-Madhuban Chowk, Panchkuian Road and Najafgarh were found as 0.01, 0.1 and 0.11 respectively. $PM_{1.0}$ to $PM_{2.5}$ ratios at different traffic corridors are presented in Figs. 7–9.

The highest $PM_{1.0}$ to $PM_{2.5}$ ratio of 1.15 during pre-scheme day was found for Najafgarh traffic corridor between 1:00 pm and 2:00 pm closely followed by value of 1.00 between 9:00 am and 10:00 am observed at Pitampura – Madhuban Chowk whereas the lowest yet very close value of 0.98 was found at Panchkuian Road traffic corridor. On the other hand, the lowest value of 0.45 was also reported at Najafgarh traffic corridor between 8:00 am – 9:00 pm followed by other two corridors reporting identical and higher value of 0.59 between 10:00 am and 11:00 am and 12:00 pm and 1:00 pm respectively (see Table 3). These higher and lower values are generally in line with traffic count and increase or decrease with that of traffic count during the particular time of day. Even though in certain cases, this trend is contributed by how the number of buses and HCVs is fluctuating. Nevertheless, these values between 0.98 and 1.15 signify the presence of very high proportion of $PM_{1.0}$ in $PM_{2.5}$ and thereby that of vehicular exhaust.

Further, if the ratios reported during the scheme days are reviewed, it is found that for all three corridors under investigation report relatively lesser values than those reported during pre-scheme days. For e.g., if maximum reported values on pre and during scheme days are compared, we see a difference of −19%; +10.11% and +5.5% at Pitampura – Madhuban Chowk, Panchkuian Road and Najafgarh Road corridors respectively. On the other side, review of minimum and lower values of the ratio yields a difference of +9.26%; +37.21% and +2.27% on the corridors in-question respectively. This pattern has also been found to be in correlation to traffic count with only exception being Pitampura – Madhuban Chowk corridor which actually has reported an increase in $PM_{1.0}$:$PM_{2.5}$ ratio even during scheme days between 8:00 am – 9:00 am. Instead of reduction in number of traffic count due to ongoing implementation of odd-even driving scheme, this corridor saw significant increase in numbers of cars and taxis, buses and HCVs which might have been so due to absence of some last moment traffic diversion on arterial roads leading to this major traffic corridor.

This relatively small reduction in ratios is attributable to various factors:

(1) ***Exemption to 2-wheelers –*** The three transport corridors in-question had about 80,307 two-wheelers plying on daily basis, constituting about 25% of total traffic count. Exemption from the purview of the odd-even driving restriction to such a large fraction of vehicles offset any major benefit. Apart from this major exemption, other exemption offered, for instance, vehicles being driven by women, school buses, government vehicles, CNG and LPG-run vehicles etc. further added to the offset.
(2) ***Duration of the restriction –*** Limiting the duration of the driving restriction only to 12 h of the day, barring Sundays from the restriction, and further implementing the policy for only a small duration of 15 days further weakened otherwise likely benefits.
(3) ***Individual and household vehicle ownership –*** The individuals either used another four-wheeler owned by them or their family members, thus swapped the vehicles for commuting; or they used two-wheelers for the 15 days of driving restriction scheme.
(4) ***Many construction sites –*** There are a number of ongoing infrastructure construction projects such as Metro rail corridors, roads, fly-overs, bridges and various other types of buildings (including major mixed-use and large buildings) have also led to this offset (as construction sites generally produce high amounts of PM).

This shows that, because vehicular emissions contribute significantly to ambient air quality, any attempt to impose or facilitate a reduction in the number of vehicles plying on the roads of Delhi city will prove helpful in reducing particulate matter (especially $PM_{1.0}$) due to vehicular pollution.

Table 4
Summary of particulate matter (PM) reduction and average ratios at various traffic corridors.

S. No.	Monitoring Station/Traffic Corridor	Average PM reduction (%)		Average of $PM_{1.0}$:$PM_{2.5}$	
		$PM_{2.5}$	$PM_{1.0}$	Pre-scheme Day	During-scheme Days
1	Pitampura – Madhuban Chowk	5.72	4.83	0.84	0.83
2	Panchkuian Road	6.11	4.70	0.8	0.7
3	Najafgarh Road	5.36	4.57	0.89	0.78

5. Conclusion

The comparative analysis of particulate matter, $PM_{2.5}$ and $PM_{1.0,}$ concentrations before and during the odd-even driving scheme implemented in Delhi city during the 15 day period of January 1–15, 2016, revealed that the observed concentrations of $PM_{2.5}$ and $PM_{1.0}$ decreased during the days of driving scheme at the three road corridors under investigation, i.e., Pitampura-Madhuban Chowk, Panchkuian Road and Najafgarh Road. The summary of average reduction of $PM_{2.5}$ and $PM_{1.0}$ during odd–even scheme days and their average ratios at various monitoring stations/traffic corridors is presented in Table 4.

It is found that the reduction of $PM_{2.5}$ is higher than that of $PM_{1.0}$ during the implementation of odd-even driving scheme. The ratios of $PM_{1.0}$:$PM_{2.5}$ at all the selected transport corridors indicate significant presence of vehicular emissions, which may be of great importance from the point of view of human health. It is also revealed from Table 4 that there is a marked reduction in $PM_{1.0}$:$PM_{2.5}$ ratio at Panchkuian Road and Najafgarh Road monitoring stations, but a very marginal decrease is observed at Pitampura–Madhuban Chowk. Considering $PM_{1.0}$ to be a good indicator for vehicular emissions at roadside locations, it is, therefore, evident that the implementation of odd-even driving rule in Delhi city during the period of January 1–15, 2016 has resulted in the reduction of vehicular emissions. The reduction in the concentration of $PM_{2.5}$ in ambient air is quite important from the point of view of having a positive impact on human respiratory health in the neighborhood and surrounding areas.

Thus, it may be concluded that the implementation of odd-even driving rule, which primarily aimed at reducing number of vehicles plying on the roads of Delhi city, has resulted in reduction of $PM_{2.5}$ and $PM_{1.0}$ in ambient air at roadside locations in Delhi city. The results of this study suggest that the adoption of such environmental measures should be considered in cities that are experiencing high levels of vehicular pollution. However, the rather small reduction in PM levels reported during the odd-even scheme days in this study also suggests that driving restriction alone cannot help long-term urban air quality improvement goals in a city like Delhi where there is substantial contribution to air quality from other sources such as construction sites, waste burning, industry, and power production. Further, it would be helpful to study the effectiveness of a full-fledged odd-even driving restriction brought about by the government with no exemption to any kind of vehicle. Additional analysis to better isolate the effects of the odd-even scheme on vehicular pollution reduction may also be pursued in the future.

A possible reason behind gross non-compliance with NAAQS values may be attributed to the rather dramatic increase in public transport buses on the city roads, which was done to ease difficulties faced by city commuters during the driving restriction days. However, in view of the fact that the heavy-duty vehicles (mainly trucks and inter-state buses) contribute very substantially to automobile emission inventories of Delhi city, and that trucks and a relatively very high number of 2-wheelers still remain on Bharat Stage (BS) III exhaust emission standards (set by Government of India), it is not surprising that air quality is not in compliance with NAAQS values. The recent announcement of Ministry of Road Transport and Highways (MoRTH) of Government of India to shift directly to Bharat Stage (BS) VI vehicle emission standards latest by April 1, 2020 is a very important step for long term reduction in vehicular emissions and improvement in ambient air quality. The new norms are expected to bring down NOx from diesel cars by 70% and in patrol cars by 25%. BS VI also makes on-board diagnostics (OBD) mandatory for all vehicles. However, timely and proper implementation of these standards needs to be ensured, calling for the development of an integrated air quality management plan to be implemented in Delhi city in near future.

Acknowledgements

The authors are thankful to Delhi Technological University for providing instrumental facilities to conduct this study.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.trd.2019.01.005.

References

Badami, M.G., 2005. Transport and urban air pollution in India. Environ. Manage. 36, 195–204.
Boogaard, H., Janssen, N.A.H., Fischer, P.H., Kos, G.P.A., Weijers, E.P., Cassee, F.R., et al., 2012. Impact of low emission zones and local traffic policies on ambient air pollution concentrations. Sci. Total Environ. 435–436, 132–140.
Brauer, M., Hoek, G., van Vliet, P., Meliefste, K., Fischer, P.H., Wigja, A., et al., 2002. Air pollution from traffic and the development of respiratory infections and asthmatic and allergic symptoms in children. Am. J. Respir. Crit. Care Med. 166, 1092–1098.
Cong, S., Zheng, S., Wang, R., 2014. Restricting driving for better traffic and clearer skies: Did it work in Beijing? Transp. Policy 32, 34–41.

CPCB - ENVIS Centre, 2016. Air Pollution in Delhi: An Analysis. Available at: http://cpcbenvis.nic.in/envis_newsletter/Air%20pollution%20in%20Delhi.pdf.
Cyrys, J., Peters, A., Soentgen, J., Wichmann, H.E., 2014. Low emission zones reduce PM10 mass concentrations. Transp. Res. D. 7, 49–64.
Davis, L., 2008. The effect of driving restrictions on air quality in Mexico City. J. Polit. Econ. 116, 38–81.
Economic Survey of Delhi, 2015. Directorate of Economics & Statistics. Govt. of NCT of Delhi.
Fensterer, V., Kuchenhoff, H., Maier, V., Wichmann, H.E., Breitner, S., Peters, A., et al., 2014. Evaluation of the impact of low emission zone and heavy traffic ban in Munich (Germany) on the reduction of PM10 in ambient air. Int. J. Environ. Res. Public Health 11, 5094–5112.
Government of India, 2017. Economic Survey 2017-2018. Ministry of Finance, Department of Economic Affairs, New Delhi, India. Available at: http://mofapp.nic.in:8080/economicsurvey/ (Accessed December 15, 2018).
Government of India, 2002. Auto Fuel Policy, Report of the Expert Committee on Auto Fuel Policy – Executive Summary. Ministry of Petroleum and Natural Gases, Government of India, New Delhi.
Gramsch, E., Le Nir, G., Araya, M., Rubio, M.A., Moreno, F., Oyola, P., 2013. Influence of large changes in public transportation (Transantiago) on the black carbon pollution near streets. Atmos. Environ. 65, 153–163.
Invernizzi, G., Ruprecht, A., Mazza, R., de Marco, C., Mocnik, G., Sioutas, C., et al., 2011. Measurement of black carbon concentration as an indicator of air quality benefits of traffic restriction policies within the ecopass zone in Milan, Italy. Atmos. Environ. 45, 3522–3527.
Lee, S.C., Cheng, Y., Ho, K.F., Cao, J.J., Louie, P.K.K., Cow, J.C., Watson, J.G., 2006. PM1.0 and PM2.5 characteristics in the roadside environment of Hong Kong. Aerosol Sci. Technol. 40, 157–165.
Pandey, A., Mishra, R.K., Shukla, A., 2016a. Study on air pollution trends (2010–2015) due to fireworks during Diwali festival in Delhi, India. Suan Sunandha Sci. Technol. J. 3 (2), 1–10.
Pandey, A., Pandey, G., Mishra, R.K., 2016b. Tailpipe emission from petrol driven passenger cars. Transp. Res. Part D 44, 14–29.
Panteliadis, P., Strak, M., Hoek, G., Wijers, E., van der Zee, S., Dijkema, M., 2014. Implementation of a low emission zone and evaluation of effects on air quality by long-term monitoring. Atmos. Environ. 86, 113–119.
Pope, C.A., Dockery, D.W., 2006. Health effects of fine particulate air pollution: lines that connect. J. Air Waste Manage Assoc. 56, 709–742.
Qadir, R., Abbaszade, G., Schnelle-Kreis, J., Chow, J., Zimmermanna, R., 2013. Concentrations and source contributions of particulate organic matter before and after implementation of a low emission zone in Munich. Ger. Environ. Pollut. 175, 158–167.
Sharma, M., Dixit, O., 2016. Comprehensive Study on Air Pollution and Green House Gases (GHGs) in Delhi. DPCC.
Singh, A.K., Gupta, H.K., Gupta, K., Singh, P., Gupta, V.B., Sharma, R.C., 2007. A comparative study of air pollution in Indian cities. Bull. Environ. Contaminat. Toxicol. 78, 411–416.
Titos, G., Lyamani, H., Drinovec, L., Olmo, F.J., Mocnik, G., 2015. Evaluation of the impact of transportation changes on air quality. Atmos. Environ. 114, 19–31.
Viard, B.V., Shihe, F., 2015. The effect of Beijing's driving restrictions on pollution and economic activity. J. Pub. Econ. 125, 98–115.
Wang, H., Fu, L., Zhou, Y., Du, X., Ge, W., 2010. Trends in vehicular emissions in China's mega cities from 1995 to 2005. Environ. Pollut. 158, 394–400.
WHO (World Health Organization), 2005. Air quality guidelines for particulate matter, Ozone, Nitrogen Dioxide and Sulphur Dioxide, Global update 2005. Geneva, 22 pages.
WHO (World Health Organization), 2014. http://www.who.int/mediacentre/news/releases/2014/air-quality/en/.
WHO (World Health Organization), 2016. http://www.who.int/phe/health_topics/outdoorair/databases/AAP_database_summary_results_2016_v02.pdf?ua=1.
World Energy Outlook Special Report, 2016. Energy and Air Pollution http://www.iea.org/publications/freepublications/publication/weo-2016-special-report-energy-and-air-pollution.html.
Yunxia, L., Zaisheng, H., Yong, L., 2015. Do driving restriction policies effectively motivate commuters to use public transportation? Energy Policy 90, 253–261.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 9, 2018, accepted December 16, 2018, date of publication January 1, 2019, date of current version January 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2890551

The Role of Big Data Predictive Analytics and Radio Frequency Identification in the Pharmaceutical Industry

MUHAMMAD NOMAN SHAFIQUE[1], MUHAMMAD MAHBOOB KHURSHID[2,3], HAJI RAHMAN[4], ASHISH KHANNA[5], AND DEEPAK GUPTA[5]

[1]Department of Business Administration, Dongbei University of Finance and Economics, Dalian 116023, China
[2]Department of Examinations, Virtual University of Pakistan, Lahore 54000, Pakistan
[3]Department of Information Systems, School of Computing, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia
[4]Department of Management Sciences, University of Buner, Buner 17290, Pakistan
[5]Maharaja Agrasen Institute of Technology, GGSIP University, New Delhi 110078, India

Corresponding author: Muhammad Mahboob Khurshid (mehboob.khursheed@vu.edu.pk)

ABSTRACT Increasing supply chain performance (SCP) has always been a common phenomenon in the history of organizations. Recent technological advancements [including, radio frequency identification (RFID) in combination with big data predictive analytics (BDPA)] have been able to significantly leverage SCP. In this respect, there is an urgent need for acceptance of RFID technology as well as BDPA to be empirically investigated in order to better understand the enhancement of SCP. Therefore, this paper aims at establishing and empirically investigating the relationship between BDPA acceptance, RFID acceptance, and SCP, respectively. The population of this paper comprises the pharmaceutical logistics industry in China. The sample was obtained from employees through the use of a simple random sampling method. Data were collected through an adapted questionnaire. SmartPLS 3.0 software was used for structural equation modeling analysis. Partial least square-structural equation modeling (PLS–SEM), the value of path coefficient, was found to be 0.700. This showed a direct positive effect of BDPA acceptance on SCP. Moreover, RFID acceptance was also tested through the ''variance accounted for'' method of mediation. Results showed a partial mediation of 57% between BDPA and SCP. Theoretically, this paper fills the knowledge gap by finding relationships between BDPA and SCP via a mediation role of RFID technology. Practically, it would be beneficial for supply chain managers in the pharmaceutical industry to adopt and implement BDPA and RFID technology so as to enhance SCP.

INDEX TERMS Big data predictive analytics, Internet of medical Things (IoMT), radio frequency identification, pharmaceutical logistic, supply chain management, supply chain performance.

I. INTRODUCTION

In this age of rapid technological advancements, Information and Communication Technologies (ICTs) such as internet, laptops, smartphones, and social media have become vital parts of everyday life making human experiences appropriately fit, smooth and systematic. These technologies connected with internet (form Internet of Things) have also made it possible to generate, collect, process, and share data in huge volumes [1]–[4]. The data is increasing continuously and rapidly in the supply chain management which demands the latest tools to utilize big data for decision-making processes. Similarly, a challenging situation has arisen for organizations in all industries to organize, utilize and predict future trends of the supply chain and to enhance performance in supply chain operations by processing varied and large amounts of data.

In the landscape of supply chain management, distributive pharmaceutical organizations in China are increasingly using their customized pharmaceutical logistics applications, in addition to collecting data on a daily basis from different channels. The data being collected on a daily basis are huge in volume, thereby generating velocity and varied natures, as well as providing significant value to decision-making processes [4]. However, since big data is generated from different sources rapidly and in large amounts, it is useless most of the time unfortunately. One of the reasons is that pharmaceutical organizations are not using this big data to predict future trends [5].

In China, barcode scanners are being used for the delivery and management of inventories in most pharmaceutical logistics organizations. These are old technologies and have limited functionalities which render them unable to track products in real-time. In addition, they have slow processing speeds. Moreover, most of the time, it has taken more than four days to deliver a product to the customer when shopping online from E-commerce websites. Also, the customers are unaware of the exact location of their products. They merely receive information regarding a few central places, which are updated through bar-code scanners. On the one hand, this creates the need for product information in real-time so as to ascertain product status while, on the other hand, pharmaceutical organizations need to track their inventory in real-time. In this regard, it is crucial for pharmaceutical organizations to be able to use advanced technologies so that they can enhance performance in their supply chain operations [6].

It is evident that not only software and soft techniques are necessary but also the hardware which is required to perform operations efficiently. In this respect, big data predictive analytics along with RFID technology are needed to be implemented in order to enhance pharmaceutical logistic performance. Therefore, the main objective to be empirically investigated in this study is the utilization of advanced technologies such as big data predictive analytics and radio frequency identification and their roles to effect supply chain performance in pharmaceutical organizations. It is a sign of the current and urgent need of the times to fill practical and operational gaps in the pharmaceutical logistics industry by investigating the role of BDPA and RFID technology in enhancing supply chain performance. The novelty of this research study lies in the fact that such relationships have never been established or empirically investigated in the pharmaceutical logistics industry. Consequently, this study opens up new horizons for academicians and supply chain managers in the field of supply chain management.

Resource-based theory (RBT) has answered this research question. It will explain this research question through establishing and empirically testing the relationship of the advanced technological resources of RFID. It will also aim to advance the technical capability of BDPA to enhance the operational process, especially SCP [5]. The advancement in technology increases the efficiency of pharmaceutical organizational processes since technology will facilitate the ability of employees to smooth the operations for productive output.

The structure of this study is systematic, with the in-depth literature review being conducted on study variables especially focusing on BDPA, RFID, and SCP. The methodology was interpreted to establish the population, instrument, validity, and reliability of data respectively. The structural model based on path analysis using PLS was conducted. At the end of the study, conclusion and discussion, as well as limitations and future directions of the research are presented.

II. LITERATURE REVIEW

A. RESOURCE-BASED THEORY (RBT)

Resource-based Theory (RBT) revolves around organizational resources and capabilities, and competitive advantage concepts. The RBT theory explains that performance level of each organization to attain competitive advantage among rivalries is different, since each organization owns diverse resources [11]–[13]. Therefore, this theory establishes a theoretical understanding on employing organizational resources and capabilities to enhance outcomes. It further elaborates that organizational resources and capabilities are considered as assets used to enable an organization to conceive and execute strategies for improving its efficiency and effectiveness. Among other organizational theories, it has been adopted widely in previous literature.

Foundations of this study are based on an organizational theory i.e. Resource-based Theory (RBT). Since, BDPA and RFID technology represent the resources and capabilities of an organization, RBT theory has been found to be suitable to examine BDPA and RFID technology roles in pharmaceutical organizations. This is for the purpose of enhancing performance in their supply chain operations. RBT has suggested that pharmaceutical organizations can utilize their resources and capabilities so as to improve their performance through competitive advantage [11]–[13]. RBT theory follows a VRIO model comprised of four characteristics, namely, Valuable (V), Rare (R), Imitable (I) and Organized (O) to enhance pharmaceutical organizational performance. Thus, if pharmaceutical logistics organizations use BDPA and RFID technology, it would be possible to improve their supply chain performance. It has also been well recognized in earlier literature that it is possible to have a relationship between RBT theory and pharmaceutical organizational performance [11], [14]–[17].

B. BIG DATA PREDICTIVE ANALYTICS (BDPA)

Big data is a considered as a technology having four characteristics which are denoted as four Vs, namely: (1) the massive Volume of data; (2) heterogeneity of data i.e. Variety; (2) rapid increase of data called Velocity; and (4) Value. The wisest decision for pharmaceutical organizations is to utilize and analyze big data in order to predict the future. However, five other characteristics have also been found in literature, specifically: Exhaustive Scope; Relational Nature; Flexibility; Resolution; and Indexical [18]. Within big data technology, big data predictive analytics can be defined as the systematic process used to capture, store, organize, use, transfer and visualize data so as to predict future trends [4], [19], [20]. Thus, pharmaceutical organizations can improve their supply chain performance through the visualization of big data referred to as big data predictive analytics [21].

Big Data analytics comprise five phases including data generation to acquisition, storage, analytics, and

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

visualization of data. In the first phase of big data analytics, data are generated mainly from the internet, in addition to audio-visual sources, remote sensing devices, and real-time applications [7]. Data from these sources are generated at a very fast pace. In the data acquisition phase, several other processes are involved such as: gathering large amounts of raw data; filtering and cleaning for future use; and sharing offline or online, with open of close access. Data are stored using big infrastructure (since it is not possible to store such a large amount of data) and utilizing different efficient storing methods to ensure better future utilization. Data analytics are then applied to extract knowledge from huge amounts of data using machine learning, computational intelligence, and different data mining techniques respectively. This extracted knowledge is used for decision-making by the stakeholders. In the last phase of big data analytics, data is visually represented in order to make sense. This graphical representation facilitates the ability of end-users to fully and easily understand the data without having through skills [7].

Big data analytics, based on types of data, have been categorized into six classes, namely: (1) Structural Data Analytics; (2) Text Analytics; (3) Web Data Analytics; (4) Multimedia Data Analytics; (5) Network Analytics and (6) Mobile Analytics [7]. As the name denotes, structural type of data is maintained by specialized systems known as Database Management Systems (DBMS) including Online Analytical Processing (OLAP). Text analytics use data extracted from social media, blogs, or emails using different techniques such as Natural Language Processing (NLP). Web data analytics involve knowledge extraction from web application sources such as web pages and cookies [7]. Multimedia data analytics involve extraction of data from audios, videos, pictures, and scanned documents. Network analytics is a type of big data analytics based on the data extracted from social networks. The last category of big data analytics is the mobile analytics; this involves: checking users' behaviors on mobile devices while playing games on mobiles; installing or uninstalling an application; and number of visits on a website [8], [9].

Big data analytics in the supply chain can be used to analyze the past, present, and future. It has three forms including: (1) Descriptive Analytics; (2) Predictive Analytics; and (3) Prescriptive Analytics. Big data predictive analytics deals with future predictions of an event which will happen or is likely to happen. It explores data patterns by using different statistical techniques, virtual simulations, and employing different programming techniques. BDPA is not merely limited to forecasting customers' tendencies towards purchasing but it can also be used throughout the organizations' operations. These include: designing and developing products; planning for demand and supply; designing supply chain networks; procurements; logistics; as well as analyzing future trends due to sudden short-term changes in the supply chain [10].

C. RADIO FREQUENCY IDENTIFICATION (RFID)

Radio frequency identification technology evolved from an electronic product code. This technology is used in supply chain management to increase the efficiency and effectiveness of network channel work in pharmaceutical organizations in order to enhance their performance [25]. Radio frequency identification technology is general in nature; hence this technology can be used in every industry and sector and is not specific to any pharmaceutical organization. This unique feature of radio frequency identification technology has provided a boost in supply chains which is equally important in all sectors [25], [27].

Radio Frequency Identification is a fast and accurate technology in comparison with doing work manually. RFID technology can automatically collect data from objects and also identify and track them accurately by the use of smart cards, optical character recognition, and barcode systems. As the name implies, RFID employs wireless technology that uses radio frequencies for transmitting the data. RFID technology comprises three main components, namely: (1) Tags; (2) Readers; and (3) a software application to interact with readers. Therefore, RFID technology is a combination of hardware and software components designed to: identify an object; collect data from tags; and track it fast and accurately [28]. The contribution of RFID in supply chain management is evident from previous literature. It is a fast technology through which entry and exit of products within an organization can be recorded, controlled, tracked, and evaluated. The tag-enabled devices are scanned through the reader with the help of special software packages to better manage the supply chain in the organization. Therefore, RFID technology, without any doubt, is helpful in increasing and improving the supply chain performance of an organization. Pharmaceutical logistics organizations are benefiting from the following: using RFID to track counterfeit medical products; provision of medical transcriptions; recording admission and discharge of patients; and tracing outdated products [29], [30]. Thus, acceptance of RFID technology is vital to be empirically investigated within pharmaceutical organizations.

The RFID tag system is the advancement of the traditional barcode system. The number of disadvantages involved with the conventional barcode system encouraged a shift towards radio frequency identification. The main problem with the traditional barcode system is that they have insufficient information and they need to scan every product manually. Moreover, once the barcode is issued it cannot be changed [27]. A barcode can be unreachable and sometimes it will misprint, which renders it unreadable. These are the significant disadvantages in the bar code system which encourage pharmaceutical organizations to move from the traditional barcode system to the systematic radio-frequency identification tag [24], [27].

D. BDPA, RFID, AND SUPPLY CHAIN PERFORMANCE

In supply chain management, big data predictive analytics have been considered as a powerful tool to build relationships between buyers and suppliers in order to boost the performance and efficiency of the supply chain [22], [23].

BDPA can be used to predict future trends and to increase productivity and operational performance in SCM using ongoing processes and existing customers. Therefore, BDPA has been considered to have a positive effect on supply chain performance [5]. In reviewing the literature, BDPA acceptance has been found to have a positive and significant impact on supply chain performance. In this study, the relationship has been analyzed through a process of structural equational modeling [2].

The use of BDPA is rapidly growing in the supply chain management field. It can be used to, specifically: enhance buyer and seller relationships; reduce costs; as well as increase operational capabilities for achieving efficiency [9]. The BDPA has a positive effect on performance as has been found from literature [8]. RBT suggested that technological resources would enable pharmaceutical organizations to enhance their performance and gain competitive advantage through SCI capabilities. Another study indicated that IT capabilities can strengthen performance through supply chain agility. Furthermore, the BDPA capability can result in competitive advantage and sustainable performance [2], [22].

In 1999, Electronic Product Code (EPC) had been introduced by Proctor & Gamble (P&G) and Gillette together with the Massachusetts Institute of Technology. The purpose of this technology is to assign automatic numbering or auto ID in order to assign a unique number to each product. It can also differentiate between different products [24]. The purpose was to "generate the intellectual fire-power to develop an open standard architecture infrastructure that would make it possible for computers to identify any object anywhere in the world" [25]. On the other hand, some scholars have said that the purpose of this technology is to move towards the Internet of things since this technology will enable pharmaceutical organizations to handle their products [26], [27].

Radio frequency identification technology minimizes human interaction because this technology does not require a direct connection for an operation like the traditional barcode system [31]. Furthermore, since RFID tags operate very fast, they can respond in milliseconds. These tags can be readable virtually, and weather conditions do not affect them. There are a number of other advantages; for example, in this technology, the product can be real-time traced at any location along the supply chain channel. Through this technology, products can communicate with each other. Moreover, this technology has been enabled to handle inventory level; products can actually inform the need for replacement of other products on their shelves [31].

The essential components of radio frequency identification technology are based on application software including server, antenna, transponders, and readers. The radio-frequency identification technology tag is a tiny microchip based on the antenna. This is used as the transponder to send the signals and communicate wirelessly with RFID readers who will receive these signals on their server [26]. In this technology, the application software enables retailers, suppliers, manufacturers, and pharmaceutical logistics providers to connect with their server to obtain real-time information about the product in the supply chain channel. This allows them to maintain the specific inventory level and save their costs [24].

Radio frequency identification technology has enabled pharmaceutical organizations to determine the flow of their products. It also calculates the total time which is required from pharmaceutical organizations to a supplier to retailer and retailer to the customer in "all the way" time. This enables greater confidence on the part of manufacturers, suppliers, retailers and customers [25]. This technology can also provide a visible flow of products in the supply chains through open communication lines [32]. Furthermore, RFID technology supports "every aspect of supply chain management, helping improve demand management, customization, and automatic replenishment of out-of-stock goods while reducing inventory and distribution costs, as well as counterfeit versions of name-brand items" [26].

RFID technology is found to be a fast-growing technology used in pharmaceutical logistics organizations to increase performance in the supply chain. The adoption of RFID technology has been well-elaborated in earlier literature [33]. This technology helps in leveraging, specifically: reduction of the operating costs; tracking inventory; reduction of lead time; increasing accuracy, efficiency, and effectiveness; as well as strengthening of buyer and seller relationships through the exchange of information, thereby increasing overall performance in the supply chain [33].

Radio-frequency identification technology is a specialized technology designed for network channel and supply-chain management. The purpose of this technology is to the track the products on a real-time basis and to maintain the level of inventory at more than the minimum level. This technology will enable pharmaceutical organizations to enhance the effectiveness of their network channel, which will lead to enhanced pharmaceutical organizational performance Thus, the implementation of radio-frequency identification technology plays a vital role in the success of network channel to reduce costs and time as well as enhancing pharmaceutical organizational performance [27].

RFID acceptance has evidenced an increase in supply chain performance, since it can enhance tracking of the inventory records, reducing operating costs and enhancing the efficiency of pharmaceutical logistics organizations. These have all led to an improvement in the supply chain performance [34]. Thus, when a pharmaceutical organization uses RFID technology in their supply chain management, the supply chain performance will then be improved automatically. In view of the stated facts, the positive relationship between RFID technology and supply chain performance has been established in this study and led to the formulation of the following hypotheses.

H1:BDPA acceptance has a positive effect on SCP.

H2:RFID acceptance has a mediating effect between BDPA acceptance and SCP.

III. METHODOLOGY

The methodology is referred to as the set of all activities performed during the research study. It is based on and includes: finding a population; extracting the sample size; determining a data collection method; selection of instrument; measurement; data analysis techniques and usage of most suitable software packages.

The target population for this study is comprised from the pharmaceutical logistics industry in China. Sampling size was selected through the application of a simple random sampling method. In this study, the sampling size was set to be 200. The employees were selected from among pharmaceutical logistics distribution organizations in China. The reason for choosing this particular population resulted from the nature of the studied variables. Distribution pharmaceutical logistics organizations are directly associated with supply chain management throughout China. They typically collect big data and use barcode reader devices (an old form of RFID technology) in the supply chain operations.

Data was collected based on the survey method through the use of an adapted questionnaire. The time of data collection spanned the period September 2017 to November 2017. A total number of 200 questionnaires were distributed among employees of pharmaceutical logistics distribution organizations in China. However, a total number of only 155 completed questionnaires out of 200 were received from respondents for carrying out further analysis. Therefore, the survey response rate was 77%.

The instrument was developed by adapting questionnaire items from previous studies. Questions about big data predictive analytics were adapted using the previous study including [35]. Moreover, items of supply chain performance were measured through eleven adapted items [36]. Additionally, items for RFID technology acceptance were measured through four adapted items [6]. All items were measured on a Likert scale; the five-point anchor was used in this study.

Advanced tools and techniques facilitated the task of the researcher to perform work efficiently in a focused and time-saving manner. In this study, different tools were used for different purposes. The online questionnaire was developed through Lime Survey; this enabled the online survey form and collection of data. It also stores data and downloads coded data into comma separated values (CSV). This CSV file can then be imported into SmartPLS 3.0 for further analysis.

SmartPLS is used to analyze data through structural equation modeling. There are two main reasons to use PLS. First, PLS can work on a relatively small sample size [37]. The logic behind working on a small sample size is that this software does not work on data normality assumptions [38], but rather, will provide precise results on a small sample size. Other tools like AMOS cannot work on a small sample size because they are required to fulfill the data normality assumptions.

The second reason to choose PLS is that this software is superior at the exploratory stage research and theory building. The relationship among BDPA, RFID, and SCP is still at the early stage. Accordingly, to test this relationship, PLS is a more useful tool, compared to other available tools [39]–[41].

In this study, Microsoft Word 2016 was used for report writing, while Endnote X8 was used for referencing in order to avoid mistakes in citations and references. Grammarly has been used to proofread and remove significant language mistakes from the manuscript.

IV. FINDINGS

A. DEMOGRAPHIC CHARACTERISTICS

The descriptive statistics of the respondents participated in the survey are projected in Table 1. According to the descriptive statistics, there were 96 males and 59 females who recorded their feedbacks. However, most of the proportion (age-wise) consisted of young participants i.e. 54.9% whereas a total number of 125 employees were having 4-9 years of experience. Further, a total number of 80 employees belonged to senior position.

TABLE 1. Respondents' demographic profile.

Category	Frequency	Percentage
Gender		
Male	96	61.9
Female	59	38.1
Age Group		
20–24 years	42	27.2
25–29 years	43	27.7
30–34 years	20	12.9
35–39 years	20	12.9
40–44 years	13	8.4
45–49 years	10	6.4
50–54 years	7	4.5
55–59 years	-	-
≥60 years	-	-
Experience		
1–3 years	15	9.7
4–6 years	75	48.3
7–9 years	50	32.3
≥10 years	15	9.7
Position		
Top Management	36	23.2
Senior Staff	80	51.6
Junior Staff	39	25.2

B. COMMON METHOD BIAS

Common method bias (CMB) is different from measurement bias. There are opportunities to use CMB when data for both independent and dependent variables have been collected at the same time from the same respondents. CMB can be checked through a variety of different techniques. However, the easiest way to find CMB is via the Bagozzi approach. This method suggests finding CMB through the correlational matrix; if the correlation between variables is less than 0.9, then there is no CMB issue in the data [42].

TABLE 2. Mean, standard deviation, correlation and square root of AVE.

Constructs	Mean	S.D.	Correlation Matrix		
			(1)	(2)	(3)
1. BDPA	1.68	0.662	**0.849**		
2. RFID	1.80	0.729	0.741**	**0.858**	
3. SCP	2.03	0.818	0.691**	0.751**	**0.830**

** significant at 0.01, bold values in diagonal are the square root of AVE.

The correlational matrix is mentioned in Table 2 below. All correlational values are less than 0.9, which shows there is no CMB in the data. In PLS, CMB also checks through VIF collinearity; values are less than 3.4 thereby showing that there is no CMB in data.

C. MEASUREMENT MODEL

A covariance-based Structural Equation Modelling (SEM) technique was used to carry out the analysis. A Partial Least Squares-SEM was used because it can run on data having a small sample size. PLS-SEM techniques were performed using SmartPLS 3.0. PLS-SEM comprises two models, namely, Measurement Model and Structural Model.

Reliability and validity are the two main components of measurement modeling. In this study, reliability was measured through Cronbach's Alpha and composite reliability [43] whereas validity was measured through content validity, convergent validity and discriminant validity respectively [43].

Reliability is referred to as the internal consistency of items [40]. In this study, both individual and composite reliability were measured. The individual reliability of items was measured through Cronbach's Alpha (α) value; the minimum accepted value is 0.6 [44]. The values of Cronbach's Alpha are shown in Table 3 below. All values were higher than the minimum acceptable value; hence, the data was found to be reliable.

Composite Reliability is the collective reliability measured through composite reliability (C.R.). The threshold value of composite reliability is 0.6 [37]. Composite reliability (C.R.) values are described in Table 3 following. All composite reliability values were greater than the minimum acceptable values thereby allowing the data to carry out further analysis.

Content validity can be checked through the logic of the content and grammar. If respondents can easily understand the meaning and structure of items, then the items have content validity. Furthermore, items must be based on some kind of logic. Content validity can be analyzed through factor loadings; if items have factor loading greater than 0.6, then they have content validity [37], [45]. Factor loading of all items have been mentioned in Table 3; all values were observed to be greater than 0.6. Thus, all items bear content validity.

Convergent validity is referred to as the extent to which items of a construct are theoretically correlated with each other. All these validities can be measured through factor loadings. If items have a factor loading of greater than 0.6, then they are considered as valid [37]. The factor loading of all items was described in Table 3; all values were greater than the cut-off value and therefore convergent validity targets were met.

TABLE 3. Factor Loading, Reliability, t-statistics and AVE.

Construct	Item Loading	t-statistics	Cronbach's Alpha	CR & AVE
BDPA				
BDPA 1	0.873	22.68	0.807	CR = 0.886
BDPA 2	0.859	30.72		AVE =
BDPA 3	0.816	20.29		0.721
RFID				
RFID 1	0.890	38.90	0.881	CR = 0.918
RFID 2	0.892	34.89		AVE =
RFID 3	0.825	18.54		0.737
RFID 4	0.824	20.78		
SCP				
SCP 1	0.885	34.82	0.954	CR = 0.961
SCP 2	0.835	25.73		AVE =
SCP 3	0.859	31.79		0.690
SCP 4	0.890	40.43		
SCP 5	0.741	13.11		
SCP 6	0.772	16.17		
SCP 7	0.717	15.12		
SCP 8	0.846	25.98		
SCP 9	0.885	41.40		
SCP 10	0.848	28.04		
SCP 11	0.835	23.35		

Discriminant validity shows how factors differ from each other. It can be measured through the square root of average variance extract (AVE). If the square root of AVE is greater than correlation values, then items have discriminant validity. The values of the square root of AVE are mentioned in Table 1, which shows that the items have discriminant validity [37], [43].

D. STRUCTURAL MODEL

The second part of PLS-SEM is to assess the structural model. The relationships between the constructs and hypotheses are tested in this part of PLS-SEM. The structural model can be assessed through path coefficient, explanatory factor (R^2) and their significant values in PLS-SEM respectively [46]. The structural model has been tested using Smart PLS 3.0. The significant values can be measured through a bootstrapping method; minimum bootstrapping is 500. In this study, 5000 bootstrapping samples were taken to enable the results to be more generalized for the study. Analysis shows that BDPA acceptance has a positive effect on supply chain performance. The direct relationship was proved by observing statistical values ($R^2 = 0.700$, t statistics = 13.28 and p-value = 0.00) during analysis. In addition, there was found to be a mediation effect of RFID acceptance between BDPA acceptance and supply chain performance. This indirect relationship was also tested statistically through the "variance accounting for" (VAF) method. The "variance accounting for" was to be 57% (ranging between 20% and 80%). The value of VAF showed the partial mediation of RFID acceptance between the relationship of BDPA acceptance and supply chain performance. The results of the direct relationship between BDPA acceptance and supply chain performance (as well as indirect, i.e. mediated relationship of

RFID acceptance between BDPA acceptance and SCP) are shown in Table 4.

TABLE 4. Path coefficients, T-statistics and sig. value.

Hypotheses	Path Coefficients	t-statistics	Sig. (p) value	Results
H_1: BDPA->SCP	0.700	13.28	0.000	**Supported**
H_2: Mediating effect	VAF = 57%	Partial Mediation		**Supported**

PLS-SEM model fit indexes have been used to find overall model fitness. In PLS-SEM, overall model fitness can be tested through the standardized root mean square residual (SRMR) and the Normed fit index (NFI). The threshold values for SRMR are less than 0.08 [47], [48]. In this study, the value of SRMR was 0.073, which was less than the threshold value of 0.08. Thus, the model was found to be fit. The second test of model fit is NFI. The acceptable range of NFI value is between 0 and 1. In this study, a VFI value of 0.891 was found, which was close to 1. Therefore, the second test also showed the goodness and, hence, overall model fit. All in all, both the model fit indexes (i.e. SRMR and NFI) fell within the acceptable range and were found to be a good fit. The path analysis of SCP, together with factor loadings and R^2 values, is mentioned in Figure 1 below.

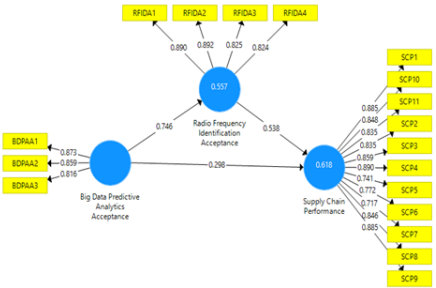


FIGURE 1. Path analysis of SCP.

V. CONCLUSION AND DISCUSSION

In this study, the relationship between BDPA acceptance, RFID acceptance and supply chain performance has been established and empirically tested in the pharmaceutical logistics industry in China. The results are aligned with earlier studies which found a significant relationship between RFID technology acceptance and supply chain performance [6]. The findings reveal that the performance of the supply chain of pharmaceutical organizations will be increased if they use and accept RFID technology.

RFID enables pharmaceutical logistics organizations to track their products in real-time and reduce misplacement of products. On the other hand, the barcode reader needs to read the barcode of products, which cannot track products in real-time. Thus, pharmaceutical logistics organizations must make a shift from barcode to RFID technology in order to enhance their performance. Customers can track their products, which will improve customer satisfaction. Similarly, the result will increase operational efficiency so as to boost pharmaceutical organizational performance.

The results also evidenced a complementary relationship between software and hardware in that software is nothing without the existence of hardware while hardware components are useless without software. RFID is the hardware technology which can provide better results if pharmaceutical organizations also concentrate on improving their soft skills of BDPA. Therefore, it is necessary that pharmaceutical organizations should have sufficient software as well as hardware resources and capabilities so as to maximum improve their operational activities.

In this study, RFID acceptance has been taken as a mediating variable in the relationship between BDPA acceptance and SCP. This mediating relationship has not been tested in previous studies. The results of this study show that if pharmaceutical organizations use BDPA and RFID together, then their performance will be increased. The empirical results of this study showed that RFID acceptance has a partially mediating effect between BDPA and SCP.

Big data is playing an increasingly significant role in the development of pharmaceutical organizations and their performance. The use of BDPA will also increase the SCP. In previous studies, the positive effect of big data on supply chain performance has been established and tested [5]. The results of this study support the previous reviews since BDPA acceptance has a positive effect on SCP.

Big data is the prominent field these days; pharmaceutical organizations have started to focus on the utilization of their data. The use of big data helps pharmaceutical organizations to learn from their previous customers' experiences and aims to fulfill the needs of their future customers. Furthermore, by the use of big data, pharmaceutical organizations can both look backward and still prepare for the future. Thus, big data learning and utilization will increase customer satisfaction in order to achieve pharmaceutical organizational performance.

This study has both theoretical and practical contributions to make in the field of SCM. This study has integrated both soft and hard resources, as well as capabilities of pharmaceutical organizations in order to boost SCP. It is envisaged that the study will enable pharmaceutical organizations to track their products, inventory levels and make their competitive strategies to achieve a sustainable competitive advantage. There is a literature gap existing in the relationship between BDPA acceptance, RFID acceptance and SCP respectively. Accordingly, this study has developed and tested this relationship, which will provide new directions for future researchers. Thus, the study is beneficial for both managers and academicians at the same time.

VI. LIMITATIONS AND FUTURE DIRECTIONS

This study has focused only on the pharmaceutical logistics industry in China. However, for generalization, this study can be tested in other developing or underdeveloped countries like India, Bangladesh, and Sri Lanka. Moreover, this study can also be implemented in other industries such as health, medical and chemical for more generalization since RFID and the big data utilization concept can be implemented in every product and manufacturing-related pharmaceutical organization to trace the products on a real-time basis to improve pharmaceutical organizational performance. This study consists of only big data and RFID technology to enhance SCP while other factors like regulatory, packaging, customer and supplier relationship, trust, and other unpredictable factors have been ignored. In future studies, these factors can be considered in order to enhance SCP since SCP is a complex phenomenon which is interlinked with the supplier, customers, and pharmaceutical organizational factors. Thus, it is highly recommended to consider these factors in future studies.

REFERENCES

[1] G. Wang, A. Gunasekaran, E. W. T. Ngai, and T. Papadopoulos, "Big data analytics in logistics and supply chain management: Certain investigations for research and applications," *Int. J. Prod. Econ.*, vol. 176, pp. 98–110, Jun. 2016.

[2] M. N. Shafique, "The role of big data predictive analytics acceptance to assimilate supply chain integration and operational performance," in *Proc. 5th ERMI Int. Conf. Arts, Educ. Social Sci.*, Cancún, Mexico, vol. 113, 2018, pp. 464–469.

[3] Z.-H. Zhou, N.-V. Chawla, Y. Jin, and G. J. Williams, "Big data opportunities and challenges: Discussions from data analytics perspectives [discussion forum]," *IEEE Comput. Intell. Mag.*, vol. 9, no. 4, pp. 62–74, Nov. 2014.

[4] L. Duan and Y. Xiong, "Big data analytics and business analytics," *J. Manage. Anal.*, vol. 2, no. 1, pp. 1–21, 2015.

[5] A. Gunasekaran *et al.*, "Big data and predictive analytics for supply chain and organizational performance," *J. Bus. Res.*, vol. 70, pp. 308–317, Jan. 2017.

[6] J. F. Kros, R. G. Richey, Jr., H. Chen, and S. S. Nadler, "Technology emergence between mandate and acceptance: An exploratory examination of RFID," *Int. J. Phys. Distrib. Logistics Manage.*, vol. 41, no. 7, pp. 697–716, 2011.

[7] M. R. Bendre and V. R. Thool, "Analytics, challenges and applications in big data environment: A survey," *J. Manage. Anal.*, vol. 3, no. 3, pp. 206–239, 2016.

[8] S. J.-F. Ren, S. F. Wamba, S. Akter, R. Dubey, and S. J. Childe, "Modelling quality dynamics, business value and firm performance in a big data analytics environment," *Int. J. Prod. Res.*, vol. 55, no. 17, pp. 5011–5026, 2017.

[9] T. Schoenherr and C. Speier-Pero, "Data science, predictive analytics, and big data in supply chain management: Current state and future potential," *J. Bus. Logistics*, vol. 36, no. 1, pp. 120–132, 2015.

[10] S. Tiwari, H. M. Wee, and Y. Daryanto, "Big data analytics in supply chain management between 2010 and 2016: Insights to industries," *Comput. Ind. Eng.*, vol. 115, pp. 319–330, Jan. 2018.

[11] J. Barney, "Firm resources and sustained competitive advantage," *J. Manage.*, vol. 17, no. 1, pp. 99–120, 1991.

[12] J. Barney, M. Wright, and D. J. Ketchen, Jr., "The resource-based view of the firm: Ten years after 1991," *J. Manage.*, vol. 27, no. 6, pp. 625–641, 2001.

[13] D. G. Sirmon, M. A. Hitt, R. D. Ireland, and B. A. Gilbert, "Resource orchestration to create competitive advantage: Breadth, depth, and life cycle effects," *J. Manage.*, vol. 37, no. 5, pp. 1390–1412, 2011.

[14] E. Brandon-Jones, B. Squire, C. W. Autry, and K. J. Petersen, "A contingent resource-based perspective of supply chain resilience and robustness," *J. Supply Chain Manage.*, vol. 50, no. 3, pp. 55–73, 2014.

[15] R. Amit and P. J. H. Schoemaker, "Strategic assets and organizational rent," *Strategic Manage. J.*, vol. 14, no. 1, pp. 33–46, 1993.

[16] T. Ravichandran, C. Lertwongsatien, and C. Lertwongsatien, "Effect of information systems resources and capabilities on firm performance: A resource-based perspective," *J. Manage. Inf. Syst.*, vol. 21, no. 4, pp. 237–276, 2005.

[17] M. Rungtusanatham, F. Salvador, C. Forza, and T. Y. Choi, "Supply-chain linkages and operational performance: A resource-based-view perspective," *Int. J. Oper. Prod. Manage.*, vol. 23, no. 9, pp. 1084–1099, 2003.

[18] M. M. Khurshid, N. H. Zakaria, and A. Rashid, "Big data value dimensions in flood disaster domain," *J. Inf. Syst. Res. Innov.*, vol. 11, no. 1, pp. 25–29, 2017.

[19] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," *MIS Quart.*, vol. 36, no. 4, pp. 1165–1188, 2012.

[20] S. Erevelles, N. Fukawa, and L. Swayne, "Big data consumer analytics and the transformation of marketing," *J. Bus. Res.*, vol. 69, no. 2, pp. 897–904, 2016.

[21] M. Barratt and A. Oke, "Antecedents of supply chain visibility in retail supply chains: A resource-based theory perspective," *J. Oper. Manage.*, vol. 25, no. 6, pp. 1217–1233, 2007.

[22] Y. Jin, M. Vonderembse, T. S. Ragu-Nathan, and J. T. Smith, "Exploring relationships among IT-enabled sharing capability, supply chain flexibility, and competitive performance," *Int. J. Prod. Econ.*, vol. 153, pp. 24–34, Jul. 2014.

[23] S. Motamarri, S. Akter, and V. Yanamandram, "Does big data analytics influence frontline employees in services marketing?" *Bus. Process Manage. J.*, vol. 23, no. 3, pp. 623–644, 2017.

[24] W. Atkinson, "Tagged: The risks and rewards of RFID technology," *Risk Manage.*, vol. 51, no. 7, pp. 12–19, 2004.

[25] N. Shister, "RFID: Taking stock of the wal-mart pilot," *World Trade*, vol. 18, no. 8, p. 38, 2005.

[26] I. Bose and R. Pal, "Auto-ID: Managing anything, anywhere, anytime in the supply chain," *Commun. ACM*, vol. 48, no. 8, pp. 100–106, 2005.

[27] M. N. Shafique, A. Nasiri, H. Rahman, and H. Ahmad, "Impact of network channel on organizational performance through radio frequency identification; Mediating role of competition," *Sarhad J. Manage. Sci.*, vol. 3, no. 1, pp. 160–171, 2017.

[28] W. Al-Khater, S. Kunhoth, and S. Al-Maadeed, "A review on radio frequency identification methods," in *Proc. 13th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2017, pp. 1751–1758.

[29] C.-P. Lee and J. P. Shim, "An exploratory study of radio frequency identification (RFID) adoption in the healthcare industry," *Eur. J. Inf. Syst.*, vol. 16, no. 6, pp. 712–724, 2007.

[30] M. N. Shafique, H. Rahman, and H. Ahmad, "The role of big data predictive analytics acceptance and radio frequency identification acceptance in supply chain performance," in *Proc. Int. Conf. Innov. Comput. Commun.* Singapore: Springer, 2019, pp. 65–72.

[31] G. Dargan, B. Johnson, M. Panchalingam, and C. Stratis, "The use of radio frequency identification as a replacement for traditional barcoding," Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep., 2004, pp. 45–877.

[32] D. Luckett, "The supply chain," *BT Technol. J.*, vol. 22, no. 3, pp. 50–55, 2004.

[33] A. White, M. Johnson, and H. Wilson, "RFID in the supply chain: Lessons from European early adopters," *Int. J. Phys. Distrib. Logistics Manage.*, vol. 38, no. 2, pp. 88–107, 2008.

[34] K. Pramatari, "Collaborative supply chain practices and evolving technological approaches," *Supply Chain Manage., Int. J.*, vol. 12, no. 3, pp. 210–220, 2007.

[35] B. T. Hazen, R. E. Overstreet, and C. G. Cegielski, "Supply chain innovation diffusion: Going beyond adoption," *Int. J. Logistics Manage.*, vol. 23, no. 1, pp. 119–134, 2012.

[36] G. D. Whitten, K. W. Green, Jr., and P. J. Zelbst, "Triple-a supply chain performance," *Int. J. Oper. Prod. Manage.*, vol. 32, no. 1, pp. 28–48, 2012.

[37] C. Fornell and D. F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," *J. Marketing Res.*, vol. 18, no. 1, pp. 39–50, 1981.

[38] C. B. Astrachan, V. K. Patel, and G. Wanzenried, "A comparative study of CB-SEM and PLS-SEM for theory development in family firm research," *J. Family Bus. Strategy*, vol. 5, no. 1, pp. 116–128, 2014.

[39] J. F. Hair, C. M. Ringle, and M. Sarstedt, "PLS-SEM: Indeed a silver bullet," *J. Marketing Theory Pract.*, vol. 19, no. 2, pp. 139–152, 2011.

[40] J. Henseler, C. M. Ringle, and R. R. Sinkovics, "The use of partial least squares path modeling in international marketing," in *Proc. New Challenges Int. Marketing, Emerald Group Publishing Ltd.*, 2009, pp. 277–319.

[41] S. Petter, D. Straub, and A. Rai, "Specifying formative constructs in information systems research," *MIS Quart.*, vol. 31, no. 4, pp. 623–656, 2007.

[42] R. P. Bagozzi, Y. Yi, and L. W. Phillips, "Assessing construct validity in organizational research," *Administ. Sci. Quart.*, vol. 36, no. 3, pp. 421–458, 1991.

[43] W. W. Chin, "The partial least squares approach to structural equation modeling," *Mod. Methods Bus. Res.*, vol. 295, no. 2, pp. 295–336, 1998.
[44] J. C. Nunnally, *Psychometric Theory*. New York, NY, USA: McGraw-Hill, 1967.
[45] J. F. Hair, Jr., W. C. Black, B. J. Babin, R. E. Anderson, and R. L. Tatham, *Multivariate Data Analysis*, vol. 3. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
[46] W. W. Chin, "How to write up and report PLS analyses," in *Handbook of Partial Least Squares*. Berlin, Germany: Springer, 2010, pp. 655–690.
[47] L.-T. Hu and P. M. Bentler, "Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification," *Psychol. Methods*, vol. 3, no. 4, p. 424, 1998.
[48] L.-T. Hu and P. M. Bentler, "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives," *Structural Equation Model., Multidisciplinary J.*, vol. 6, no. 1, pp. 1–55, 1999.



MUHAMMAD NOMAN SHAFIQUE received the M.Phil. degree in business administration from Pakistan. He is currently pursuing the Ph.D. degree in business administration from the Dongbei University of Finance and Economics, China. He has completed many research certifications and has published few research papers in different journals. He is learning new research techniques and implementing his studies.



MUHAMMAD MAHBOOB KHURSHID received the M.Sc. degree in project management from the COMSATS Institute of Information Technology, Islamabad. He is currently pursuing the Ph.D. degree in information systems with Universiti Teknologi Malaysia. He has 10 years of professional experience carrying out operations regarding examinations with the Virtual University of Pakistan, where he is currently an Assistant Controller Examinations with the Department of Examinations.



HAJI RAHMAN received the M.S. degree in engineering management from the University of Engineering and Technology, Taxila, Pakistan, and the Ph.D. degree in management sciences (HRM) from Preston University, Islamabad, on HEC scholarships. He was a Faculty Member for more than nine years with the Ghulam Ishaq Khan Institute with the Best Result Producing Teacher Awards. He was an Assistant Professor with Preston University, Islamabad campus, for two years. He also delivered trainings modules with the CDA Training Academy and the Pakistan Manpower Institute, Islamabad. He is a Visiting Faculty Member with the IIUI, Islamabad, and the Federal Urdu University of Arts Sciences and Technology, Islamabad. He is currently an Assistant Professor with the Department of Management Science, University of Buner, Khyber Pakhtunkhwa, where he also holds the additional charge as the Director Quality Enhancement Cell. He has authored a number of international and national research papers. He has presented his research articles in a number of international research conferences. He received the Punctuality and Efficiency Awards from Preston University, in Fall Semester 2015, Spring Semester 2016, Fall 2016, and Spring 2017.



ASHISH KHANNA received the B.Tech. and M.Tech. degrees from GGSIPU, New Delhi, India, in 2004 and 2009, respectively, and the Ph.D. degree from the National Institute of Technology, Kurukshetra, in 2017. He has around 15 years of expertise in Teaching, Entrepreneurship, and Research and Development, with the specialization in Computer Science Engineering Subjects. He is a Postdoctoral Fellow with the Internet of Things Lab, Inatel, Brazil. He is currently working as an Assistant Professor with the Department of Computer Science and Engineering, Maharaja Agrasen Institute of Technology, under GGSIPU, New Delhi. He has around 44 research papers and book chapters in reputed SCI, Scopus journals, conferences, and reputed book series, including 21 papers in SCI indexed Journals of Springer, Elsevier, IEEE, and Wiley Journals. In addition, he has authored, edited, and editing 17 books for some of the top publishers, such as Springer, Wiley, Elsevier, and many more. His research interest includes distributed systems and its variants (MANET, FANET, VANET, IoT), machine learning, evolutionary computing, and many more. Furthermore, he has served the research field as a Faculty Resource Person/Session Chair/Reviewer/TPC member in various conferences and journals. He received the Best Paper Award in a Springer International Conference. He has also been invited as the Keynote Speaker. He is Convener and Organizer of the ICICC-2018 and 2019 Springer Conference. He played a key role in the origination of a reputed publishing house (Bhavya Books), having 250 solution books and around 60 text books. He is also the originator of a research unit under the banner of Universal Innovator. He has designed question papers for multiple government organizations, and played a key role in promoting Smart India Hackathon at MAIT. He also serves as the Guest Editor in various reputed International journals of Inderscience, IGI Global, Bentham science, and many more. He is currently editing some books of Elsevier, Springer, and Wiley, which will be published soon. His Cumulative Impact Factor is above 50.



DEEPAK GUPTA received the Ph.D. degree from Dr. A.P.J. Abdul Kalam Technical University, in 2017, and the Master of Engineering from the Delhi College of Engineering, in 2010. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Maharaja Agrasen Institute of Technology, GGSIP University, New Delhi, India. He has published many scientific papers in SCI journals, such as JoCS, IAJIT, and FGCS. In addition, he has authored/edited over 25 books (Elsevier, Springer, IOS Press, and Katson) and six journal special issues. His research area includes software usability, human–computer interaction, nature-inspired computing, machine learning, and soft computing. He is a Convener and an Organizer of ICICC-2018 springer conference. He has also started a research unit under the banner of "Universal Innovator." He is also associated with various professional bodies, such as ISTE, IAENG, IACSIT, SCIEI, ICSES, UACEE, the Internet Society, SMEI, IAOP, and IAOIP.

Hindawi
International Journal of Antennas and Propagation
Volume 2019, Article ID 6087039, 13 pages
https://doi.org/10.1155/2019/6087039

Hindawi

Research Article

UWB Metamaterial-Loaded Antenna for C-Band Applications

Parul Dawar,[1] **N. S. Raghava**,[2] **and Asok De**[2]

[1]*Guru Tegh Bahadur Institute of Technology, GGSIPU, Department of ECE, Delhi, India*
[2]*Delhi Technological University, Department of ECE, Delhi, India*

Correspondence should be addressed to Parul Dawar; paru.dawar@gmail.com

Received 17 January 2018; Revised 11 June 2018; Accepted 16 October 2018; Published 6 January 2019

Academic Editor: Luciano Tarricone

Copyright © 2019 Parul Dawar et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A novel metamaterial-inspired patch antenna is proposed, wherein a 2-segment SRR Labyrinth metamaterial is embedded inside the antenna substrate. It is observed that upon incorporation, the bandwidth widens to around 600% and VSWR improves by approx. 1.5% and the antenna is miniaturized by 400%. The Nicolson-Ross-Weir (NRW) method has been used to retrieve the material parameters from transmission and reflection coefficients.

1. Introduction

Nowadays, the trend is developing devices for wireless communication systems which have inherent high bandwidth, gain, and multiresonance. Rectangular microstrip patch antenna, i.e., RMPA, is not a good candidate as it has a narrow bandwidth, in sufficient gain and directive properties. Thus, various optimization techniques have been introduced to make patch antenna a successful candidate in wireless communication applications.

With the development of metamaterials came the concept of attaining altered behaviour of electromagnetic radiations, i.e., μ, ε, and η, thus providing a better solution for attaining enhanced antenna's performance. "Metamaterials" (MTMs) are engineered to modify the bulk permeability and/or permittivity of the medium [1]. It is realized by placing, periodically, structures that alter the material parameters, with elements of size less than the wavelength of the incoming electromagnetic wave. It results in "meta," i.e., "altered," behaviour or behaviour unattainable by natural materials. In this paper, a novel antenna structure is proposed with enhanced performance parameters and miniaturization using metamaterial.

2. Metamaterial Design

The study of microstrip patch antennas has made progress in recent years. Compared with conventional antennas, microstrip patch antennas have more advantages and better prospects. They are lighter in weight and have low volume, low cost, low profile, small dimension, ease of fabrication, and conformity. Moreover, the microstrip patch antennas can provide dual and circular polarizations, dual-frequency operation, frequency agility, broad bandwidth, feed line flexibility, and beam scanning omnidirectional patterning. In literature, the following structures have been discussed to improve the bandwidth of the patch antenna:

(1) Defected ground structure

(2) Electromagnetic band-gap structures

(3) Metamaterials

The basic element of DGS is a resonant gap or slot in the ground metal. It helps in the reduction of surface waves, but the disadvantage of using DGS is reduction in antenna efficiency due to back lobe formation. EBGs are a periodic arrangement of dielectric or metallic elements in one-, two-, or three-dimensional manner. By introducing the structure, the surface waves are reduced as the ground/patch is loaded with inductance and capacitance, but the disadvantage of using EBG is that the area of the antenna is increased. By using metamaterials, the antenna's bandwidth is enhanced and also multiresonance may be observed by studying its response to the incident EM wave. The behaviour of the metamaterial unit cell can be studied in respect of

polarization when the incident electromagnetic wave is perpendicular to it, which is depicted by a resonating curve between its permittivity and permeability with respect to frequency. In negative index materials, phase and energy velocities follow the left hand rule. In order to observe the negative group and phase velocities, a positive index material slab is embedded with negative index material in the middle as shown in Figure 1.

The group velocity follows the right hand rule and propagates towards the right irrespective of the medium as shown in Figure 1. When ϵ and μ both are negative, phase velocity follows the left hand rule but the direction of propagation is reversed. This leads to the formation of "backward wave" in materials with negative refraction. However, if either ϵ or μ is negative, then evanescent waves occur. Many structures have been analyzed in literature [3–10] to establish this theory of left-handedness and composite material concept. This analysis of various 2D and 3D structures is based on several factors like spacing, size, arrangement, shapes, compositions, density, or materials used. In Figure 2, a generic view of metamaterials made with periodic inclusions resulting in an effective bulk permittivity and permeability is shown.

3. Evolution of the Shape of the Proposed Metamaterial

In literature [11, 12], the geometry of the SRR consists of two square shape rings with their slit displaced by 180° or 2π as shown in Figure 3.

The split or gap (g) of SR is 0.8 mm wide, the ring width (t) is 35 microns, and substrate thickness (h) is 1.6 mm. The spacing (s) between the nth and $(n+1)$th turn is 0.4 mm. The SRR is printed on a FR4 substrate having relative permittivity, εr, as 4.4.

A unit cell is designed with the above dimensions, and the simulation result is shown in Figure 4.

By introducing one split each in respective arms, capacitances in gaps are introduced. As per the equivalent circuit, the more the gaps, the more will be the Cg introduced in parallel, which means reduced effective capacitance. Thus, the resonant frequency increases. Therefore, resonant frequency is calculated to be 2.08 GHz.

If the square ring is single, i.e., only the outer ring is considered with a single split, then the resonant frequency increases. This is because the capacitance reduces as there is no coupling between the rings. The resonant frequency is calculated to be 2.83 GHz.

Therefore, the physical explanation is in coherence with the theoretical explanation of the change in resonant frequency of different configurations of square SRR. The above unit cells are designed, and the simulated result is shown in Figure 5.

Thus, an analysis can be made on the two different shapes of the SRR, i.e., circular and square, with single split and double split each keeping the outer perimeter same. A curve in Figure 6 summarizes the inferences.

The above curve has been proved using physical analysis. Also, the reason for the lower resonant frequency in square

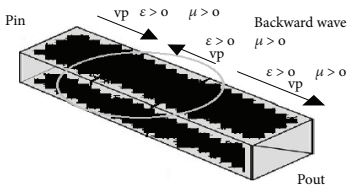


Figure 1: Antiparallel phase and group velocity if waveguide is made of metamaterial.

shape SRR than in circular SRR can be justified from the current distribution as shown in Figure 7.

The square SRR has greater current density than the circular one. It means more inductance, hence the lower resonant frequency. Thus, different shapes can be derived from these basic two shapes, i.e., square SRR and circular SRR. The advantages of the changing shape of the split-ring resonator from primitive circular shape to square are:

(1) Better degree of control: a circular SRR has a radius and width of the ring as the parameters that can be varied to control the constitutive material parameters of the medium. However, using square shape SRR, an extra variable parameter is available. Thus, a better degree of control is achievable using length, breadth, and width of the ring in square SRR.

(2) Better candidate as metamaterial: using a square shape SRR, the resonant frequency of the unit cell is reduced. Thus, (a/λ), which is the ratio of the unit cell dimension to the wavelength of operation, increases. This makes the metamaterial a better candidate over circular SRR.

The epsilon-negative group (i.e., negative permittivity) of metamaterials, shown in Figure 2(a), provided an appreciable increase in bandwidth, which are analogous to a medium consisting of thin wires arranged in a periodic array [12]. It is also observed that the permittivity ϵ varies with frequency ω of operation of the unit cell as presented in Equation (1):

$$\varepsilon(\omega) = 1 - \frac{\omega p^2}{\omega(\omega + i\gamma)}, \quad (1)$$

where ω_p is the plasma frequency and γ is a damping term.

In Equation (1), if the plasma frequency is less than the operating frequency, then the permittivity becomes negative.

In this paper, the mu-negative group (i.e., negative permeability) metamaterial shown in Figure 2(b) is explored as a 2-segment labyrinth metamaterial. If the rod and ring materials are combined, a left-handed material [13] is formed having simultaneously negative ϵ and μ. The reason is that the wire strips affect the ϵ and the split-ring resonators (SRRs) alter the μ of the medium. Thus, a frequency-dependent negative material with both the parameters being negative is obtained.

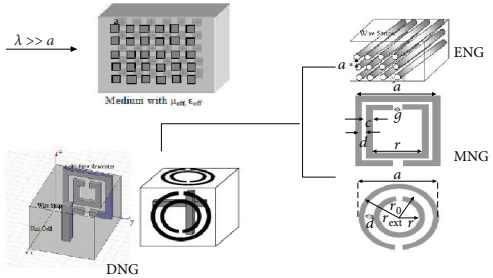


Figure 2: Generic view of a host medium with periodically placed structures constituting a metamaterial, ENG (epsilon-negative group), MNG (mu-negative group), and DNG (double-negative group).

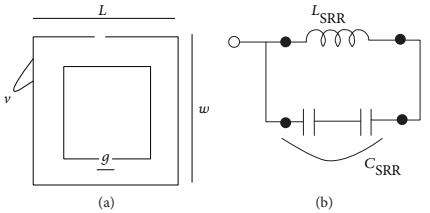


Figure 3: Geometry of square SRR and its equivalent circuit.

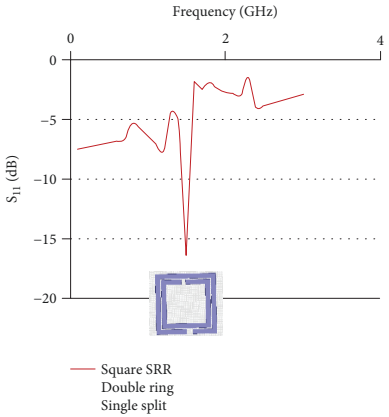


Figure 4: Unit cell analysis of square SRR.

4. 2-Segment SRR Metamaterial

The evolution of square SRR into double-split SRR has been explained in the previous section. The 2-segment labyrinth metamaterial design is as shown in Figure 8. The values of the design parameters are conductivity ρ_s of 0.034×10^{-3}

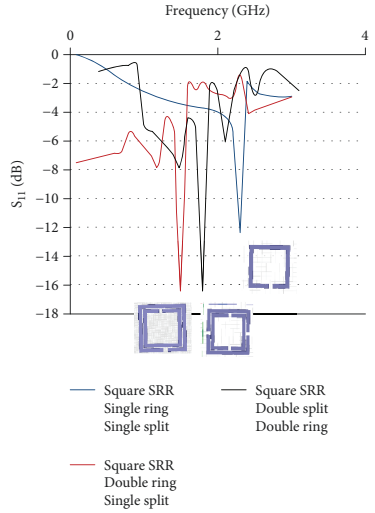


Figure 5: Unit cell analysis of different configurations of square SRR.

Ωm, side length l of the external ring of 10.2 mm, width of the strips u of 1.8 mm, separation between two adjacent strips s of 0.4 mm, gap width g of 0.8 mm, and thickness t of the conducting metallic inclusions of 0.16 μm.

This metamaterial is embedded in the middle of the antenna's substrate such that the center of patch coincides with the center of the metamaterial. The structure is then simulated using HFSS software. The top view of the 2-segment SRR metamaterial RMPA is shown in Figure 9.

4.1. Mathematical Analysis of the 2-Segment SRR Metamaterial. A 2-segment labyrinth metamaterial discussed in the previous section in Figure 8 is analyzed theoretically, and the results are compared with mathematical analysis

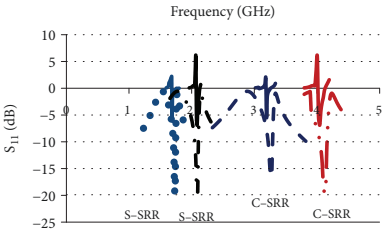


Figure 6: Comparison of circular (C) and square SRR (S) resonant frequency.

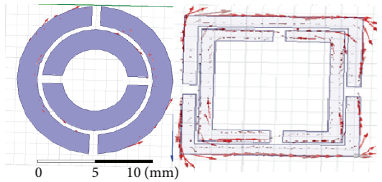


Figure 7: Current distribution in circular and square SRR.

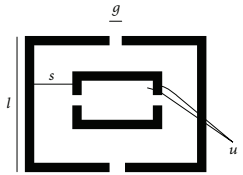


Figure 8: 2-Segment SRR metamaterial.

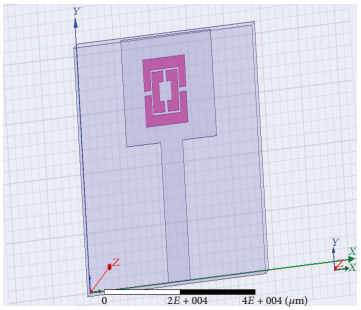


Figure 9: Top view of 2-segment SRR metamaterial RMPA.

using an equivalent circuit in this section. The Nicholson-Ross-Weir (NRW) method [14, 15] is used to retrieve or extract the parameters to observe the permeability region of SRR. The metamaterial in the dielectric substrate is bounded

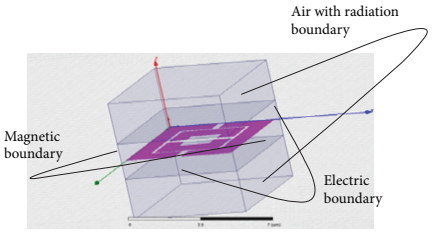


Figure 10: 2-Segment MTM: unit cell for parameter extraction.

by a box on either side as shown in Figure 10 and has air and radiation boundary on the top and bottom.

Using the NRW equation ((2)), the resonance in permeability is obtained as shown in Figure 6. The variables of Equation (2) are defined, and values of μ_r and ϵ_r are derived through the HFSS software.

$$\begin{bmatrix} \mu_r = \dfrac{2}{jK_0 d} \cdot \dfrac{1 - V_2}{1 + V_2} \\ \varepsilon_r = \dfrac{2}{jK_0 d} \cdot \dfrac{1 - V_1}{1 + V_2} \end{bmatrix}, \tag{2}$$

where k_0 is a wave number, d is the thickness of the substrate, and V_1 and V_2 are the composite terms to represent the addition and subtraction of scattering parameters and are given by Equation (3):

$$\begin{bmatrix} V_1 = S_{21} - S_{11} \\ V_2 = S_{21} + S_{11} \end{bmatrix} \tag{3}$$

Using Equations (2) and (3), the resonance in permeability is obtained in Figure 11.

It is observed in Figure 11 that the magnetic resonance is obtained at 4.2 GHz. The resonance in permeability is because a 2-segment SRR MTM has magnetic properties due to internal inductances and capacitances. Thus, the 2-segment labyrinth metamaterial behaves as a mu-negative group (MNG) having negative permeability over some frequency region [2].

The equivalent circuit of the above unit cell can be drawn as shown in Figure 12. It can be made as a combination of parallel RC and series of RL components using transmission line theory. The inductance L is taken for unit length of the loop, and capacitance C can be taken for the gap in the loop. The series resistance R_1 describes the losses in the conductor and the shunt resistance R_2 for the losses in the dielectric.

The inductance per unit length of the loop (L) is given by Equation (4) [16]:

$$L = \frac{\mu 0}{2} \frac{l\text{avg}}{4} 4.86 \left[\ln\left(\frac{0.98}{\rho}\right) + 1.84\rho \right], \tag{4}$$

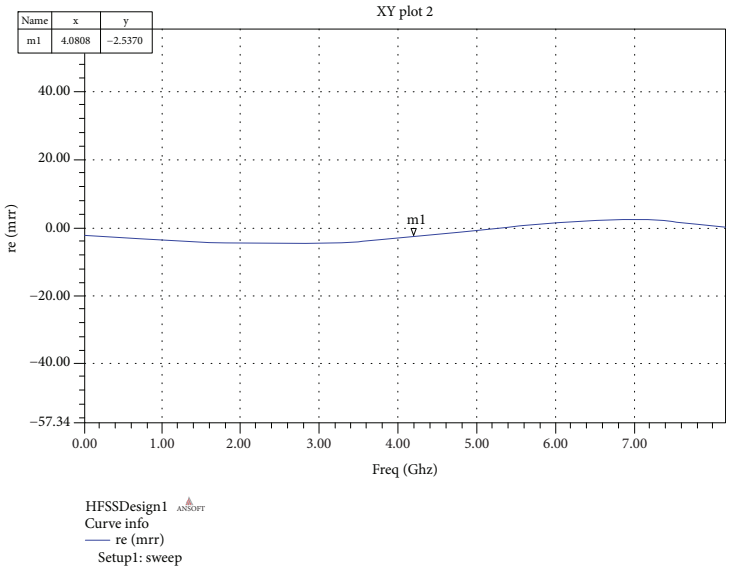


Figure 11: Resonance in permeability.

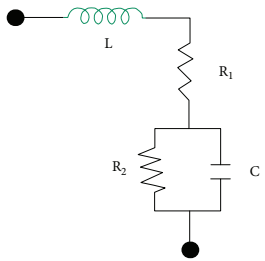


Figure 12: Equivalent circuit of 2-segment SRR metamaterial.

where μ_0 is the vacuum permeability and l_{avg} is the average strip length and is calculated over all the rings using Equation 5:

$$l\text{avg} = 4[l - (N-1)(u+s)], \tag{5}$$

where ρ is the filling ratio of the SRR between the cross-links.

Using Equation (5), l_{avg} is calculated as 36 mm. Therefore, substituting it in Equation (4), inductance L is obtained as 3×10^{-6} H.

Similarly, the capacitance of the gap (C) is given by Equation (6):

$$C = \frac{N-1}{2}[2l - (2N-1)(w+s)]C_0, \tag{6}$$

where C_0 is the per-unit-length capacitance between two parallel strips and is given by Equation (7):

$$C0 = \varepsilon 0 \varepsilon r^{\text{sub}} \frac{K\left(\sqrt{1-k^2}\right)}{K(k)}, \tag{7}$$

$\varepsilon r^{\text{sub}}$ is the effective relative permittivity related to the dielectric filling of the substrate and is given by Equation (8):

$$\varepsilon r^{\text{sub}} = 1 + \frac{2}{\pi} \text{artg}\left[\frac{h}{2\pi(w+s)}\right](\varepsilon r - 1), \tag{8}$$

where

u is the width of the segment of MTM
s is the separation between the 2 segments of MTM
h is height of substrate of MTM
$\varepsilon 0$ is the vacuum permittivity

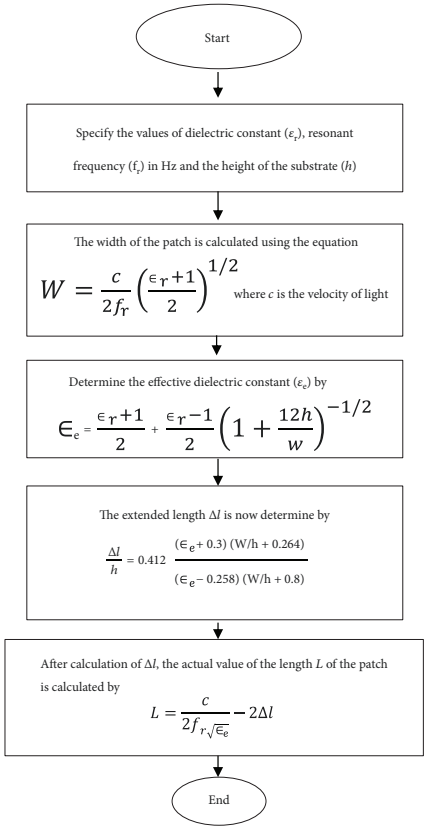


Figure 13: Design methodology for patch antenna.

K is the complete elliptic integral of the first kind and is given by Equation (9):

$$k = \frac{s}{s+2u}. \tag{9}$$

Using the above equations, C is 14.15×10^{-7} F.

The resultant impedance of the circuit in Figure 12 is given by Equation (10):

$$Z = sL + R1 + \frac{1}{sC+R2}, \tag{10}$$

where $s = j\omega$ and $\omega = 2\pi f$ where f is the resonant frequency.

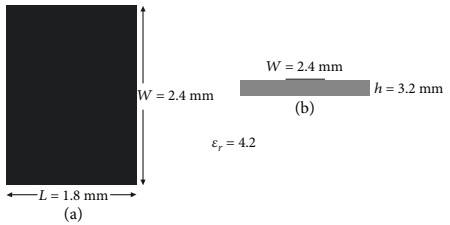


Figure 14: (a) Top view and (b) side view of the designed rectangular microstrip patch antenna.

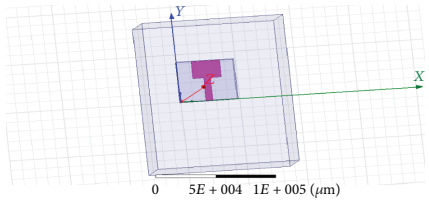


Figure 15: Rectangular microstrip patch antenna.

The frequency domain equivalent of Z is obtained by substituting $j\omega$ in place of s in Equation (10) and is given by Equation (11):

$$Z = j\omega L + R1 + \frac{1}{j\omega C + R2}. \tag{11}$$

At resonance, the imaginary part of Z is zero. So, Equation (11) becomes:

$$j\omega L + \frac{j\omega C}{-\omega^2 C^2 - R2^2} = 0. \tag{12}$$

Neglecting $R2$, the value of ω is given by Equation (13):

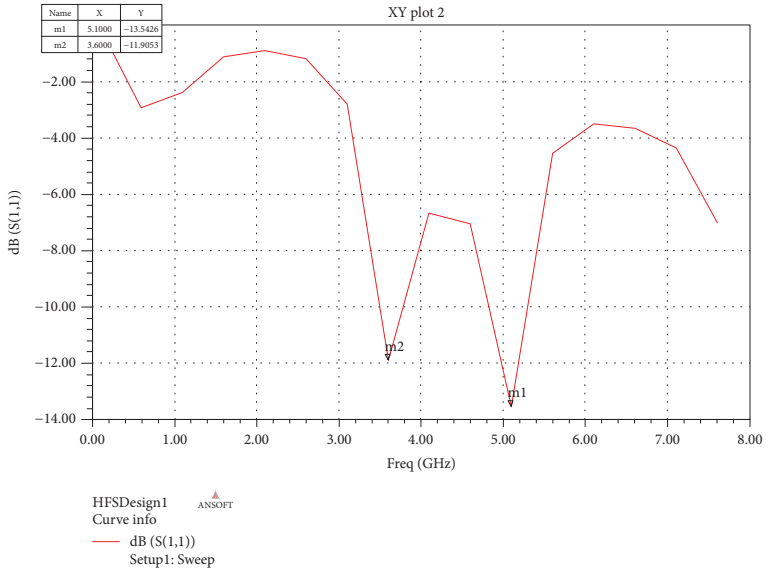
$$\omega = \sqrt{\frac{1}{L \cdot C}}. \tag{13}$$

Substituting the values of L and C in Equation (13), the resonant frequency is given as 4.08 GHz. Thus, around 1% error is seen between the simulation and the theoretical study of 2-segment SRR.

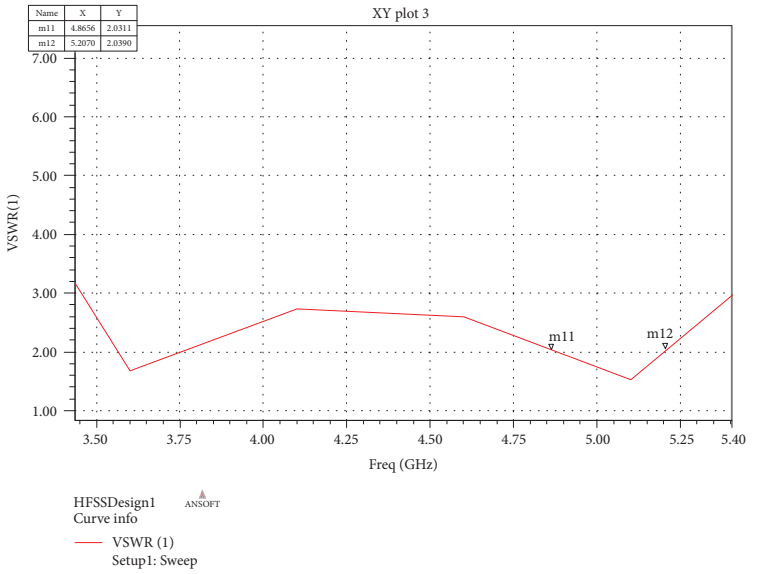
5. Antenna Design

A rectangular microstrip patch antenna is designed for 4.12 GHz as per the methodology shown in Figure 13 with design parameters as shown in Figure 14; the dielectric constant of the substrate is 4.2, FR4, and the thickness of the substrate is 3.2 mm.

The rectangular patch is etched on one side of the substrate, and the other side of the substrate acts as the ground plane. This whole antenna is excited by a quarter

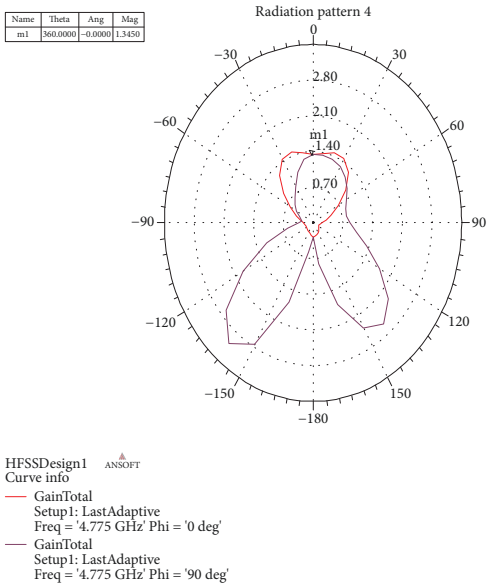


(a)



(b)

Figure 16: Continued.



(c)

Figure 16: Simulation results of antenna for 2-segment labyrinth: (a) return loss, (b) VSWR, and (c) radiation pattern in E and H planes of RMPA.

Table 1: Various parameters of RMPA at 4.775 GHz.

Parameters	RMPA
S_{11} (dB)	−13.5
VSWR	1.49
Bandwidth (GHz)	0.4
E plane gain in dB	5.8
Peak directivity (dB)	4.6

wave-transformed matched microstrip feed as shown in Figure 15.

The patch is simulated using Ansoft HFSS software, and the simulation results of the different performance parameters of antenna are shown in Figure 16.

The above simulation results are tabulated in Table 1.

For this antenna, the return loss is −13.5 with VSWR of 1.49. This shows that the impedance matching of the antenna is within the limits. The bandwidth of the antenna is 0.4 GHz. To improve the performance parameters of the above antenna, the structure of this conventional rectangular patch is modified by embedding a 2-segment labyrinth metamaterial in the middle of the substrate. This metamaterial is embedded in the middle of the antenna's substrate such that the center of patch coincides with the center of the metamaterial. The structure is then simulated using HFSS software.

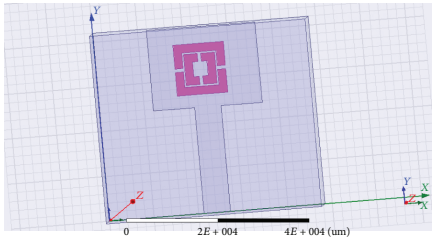


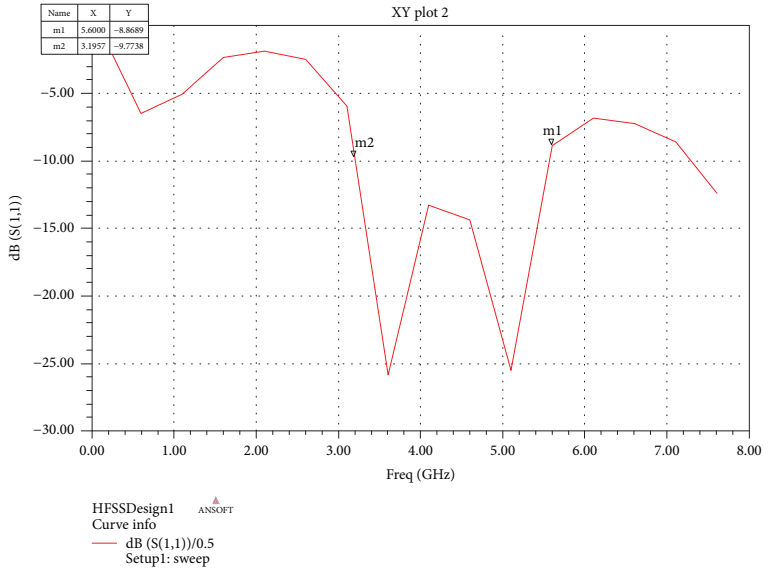
Figure 17: Top view of 2-segment SRR metamaterial RMPA.

The top view of the 2-segment SRR metamaterial RMPA is shown in Figure 17.

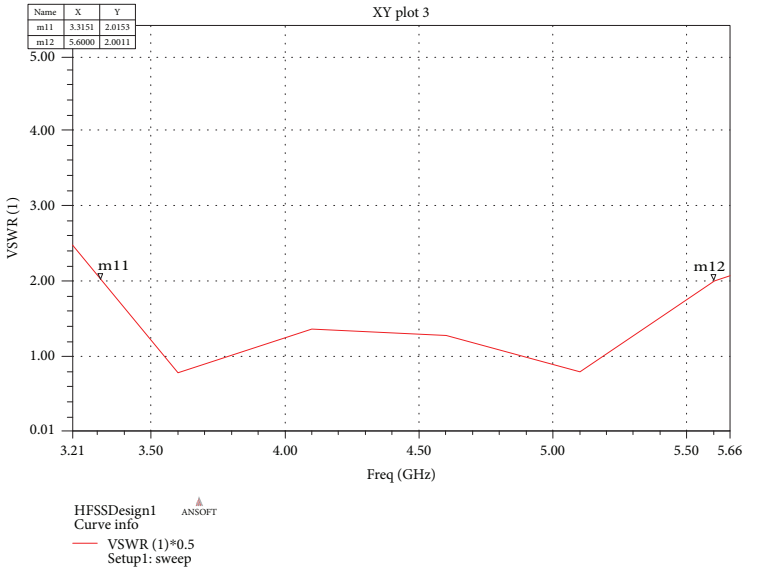
The study of performance parameters is shown in the simulation of this designed 2-segment SRR metamaterial RMPA shown in Figure 18.

The performance parameters of the MTM RMPA are shown in Table 2.

By comparing Tables 1 and 2, it can be observed that by embedding a 2-segment labyrinth metamaterial in the middle of the substrate of the patch antenna, the bandwidth increases by 6 times, in the GHz range. This lowers the effective permittivity of the antenna's substrate, which

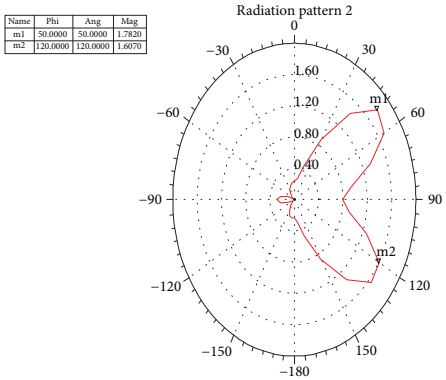


(a)



(b)

Figure 18: Continued.



(c)

Figure 18: Simulation results for 2-segment SRR metamaterial RMPA: (a) return loss, (b) VSWR, and (c) radiation pattern in E and H planes.

Table 2: Various parameters of labyrinth MTM embedded in RMPA substrate.

Antenna/parameters	RMPA with labyrinth MTM
S_{11} (dB)	−25
VSWR	1.26
Bandwidth	2.3
E plane gain dB	3.18



(a) (b)

Figure 19: Fabricated antenna: microstrip patch antenna: (a) top view and (b) bottom view.

leads to a spreading of the electric field. This in turn leads to broadening of the bandwidth. Improvement in VSWR by 1.5% indicates better impedance matching of the antenna when it is loaded with the metamaterial than conventional patch antenna. However, the trade-off is reduction in gain by 1.5%. This is because the product of gain and bandwidth



Figure 20: Keysight Fieldfox RF Network Analyzer.

is unity. So, in this structure, gain decreases due to an increase in bandwidth. It can be seen that the miniaturization of nearly $1/4^{th}$ of the conventional patch is obtained as secondary resonance is observed at 3.19 GHz.

The proposed antenna operates at the C-band, which is used in satellite applications and is more compact and broadband than the actual antenna that would be designed at C-band frequencies.

6. Experimental Validation of Simulated Results

This rectangular microstrip patch antenna has been fabricated on FR4 substrate. The fabricated antenna is shown in Figure 19.

Using the Keysight Fieldfox RF Network Analyzer as shown in Figure 20, the simulation and fabrication results are compared in Figure 21.

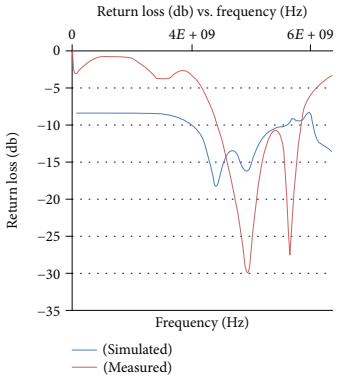


Figure 21: Return loss of patch antenna.



Figure 22: AMITECH gain measurement set-up.

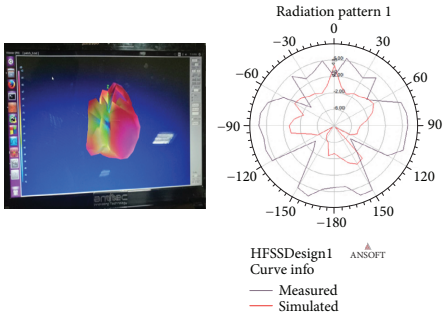


Figure 23: Radiation pattern measurement of patch antenna.

Using AMITECH, the gain measurement set-up is as shown in Figure 22.

Using the gain measurement set-up, a radiation pattern is obtained as in Figure 23.

The patch antenna is fabricated by embedding a 2-segment SRR metamaterial in the middle of the substrate as shown in Figure 24.

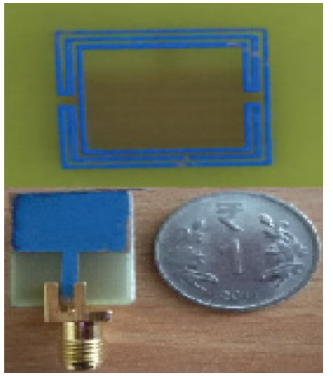


Figure 24: Fabricated antennas: microstrip patch antenna.

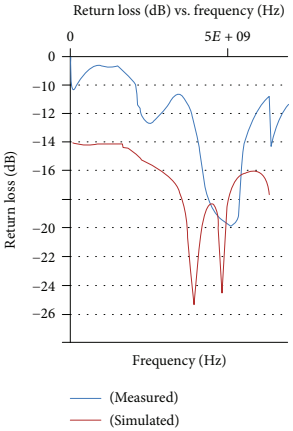


Figure 25: Return loss of metamaterial patch antenna.

Using the Keysight Fieldfox RF Network Analyzer as shown in Figure 20, the simulation and fabrication results are compared in Figure 25.

Thus, "fabricated and simulated antennas" help in achieving miniaturization and bandwidth enhancement which is in coherence with the proposed antenna configurations with nearly 5% error. It is normal to have some differences between simulated and measured results because of the following reasons:

(a) Fabrication tolerances (e.g., SMA soldering) and measurement error/limitations

(b) In simulation, when the substrate is selected, it shows a perfect material; however, variations in

Table 3: Comparison results of 2-segment SRR MTM in RMPA substrate with literature.

Reference no.	% Improvement in bandwidth of MPA	% Miniaturization
[17]	−5%	4%
[18]	−2%	24%
[19]	43.6%	41%
[20]	78.5%	4%
Proposed configuration	600%	25%

thickness & dielectric constant exist in commercially available material (especially the latter which has more influence).

(c) In actual measurement, i.e., if measurement is done inside a room/lab, there may be a lot of metallic objects which will highly affect results due to limited surrounding reflection phenomena. The influence could be on a worst side or best side; i.e., depending upon reflected signal, it can be either additive or subtractive.

In literature [17–20], using 2-segment SRRs, the bandwidth of the patch antenna is reduced, but 24% miniaturization is obtained. However, using 2-segment SRR in the middle of the antenna's substrate, an increase in the bandwidth by 600% and miniaturization of nearly 1/4th, i.e., 25% of the conventional patch, is obtained as secondary resonance is observed at 3.19 GHz. Thus, the proposed configuration is better than the previous configurations in literature. The findings can be tabulated as shown in Table 3.

7. Conclusion

In this paper, a proposed novel microstrip antenna attains miniaturization by nearly 400% and maximization in bandwidth of nearly 600% with metamaterial embedded in the middle of the substrate of the patch antenna. This is a result of negative permeability of the metamaterial which reduces the effective medium below the patch.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J. B. Pendry, A. J. Holden, D. J. Robbins, and W. J. Stewart, "Magnetism from conductors and enhanced nonlinear phenomena," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 11, pp. 2075–2084, 1999.

[2] B. Ferguson and X.-C. Zhang, "Materials for terahertz science and technology," *Nature Materials*, vol. 1, no. 1, pp. 26–33, 2002.

[3] D. R. Chowdhury, R. Singh, M. Reiten, J. Zhou, A. J. Taylor, and J. F. O'Hara, "Tailored resonator coupling for modifying the terahertz metamaterial response," *Optics Express*, vol. 19, no. 11, pp. 10679–10685, 2011.

[4] C. Caloz and T. Itoh, "Metamaterials for high-frequency electronics," *Proceedings of the IEEE*, vol. 93, no. 10, pp. 1744–1752, 2005.

[5] R. P. Clayton, *Inductance: Loop and Partial*, Wiley, Hoboken, NJ, USA, 2009.

[6] R. E. Collins, *Foundations for Microwave Engineering*, McGraw-Hill Book Company, New York, NY, 1966.

[7] D. Hyun Lee, A. Chauraya, Y. Vardaxoglou, and W. S. Park, "A compact and low-profile tunable loop antenna integrated with inductors," *IEEE Antennas and Wireless Propagation Letters*, vol. 7, pp. 621–624, 2008.

[8] D. Psychoudakis, Y. H. Koh, J. L. Volakis, and J. H. Halloran, "Design method for aperture-coupled microstrip patch antennas on textured dielectric substrates," *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 10, pp. 2763–2765, 2004.

[9] D. R. Smith, S. Schultz, P. Markoš, and C. M. Soukoulis, "Determination of effective permittivity and permeability of metamaterials from reflection and transmission coefficients," *Physical Review B*, vol. 65, no. 19, article 195104, 2002.

[10] D. R. Smith, J. B. Pendry, and M. C. Wiltshire, "Metamaterials and negative refractive index," *Science*, vol. 305, no. 5685, pp. 788–792, 2004.

[11] C. T. Chan, S. Datta, Q. L. Yu, M. Sigalas, K. M. Ho, and C. M. Soukoulis, "New structures and algorithms for photonic band gaps," *Physica A: Statistical Mechanics and its Applications*, vol. 211, no. 4, pp. 411–419, 1994.

[12] C. Caloz and T. Itoh, *Electromagnetic Metamaterials: Transmission Line Theory and Techniques*, Wiley, 2006.

[13] H. Schantz, *The Art and Science of Ultra-Wide Band Antennas*, Artech House, 2005.

[14] B. Danana, B. Choudhury, and R. M. Jha, "Design of high gain microstrip antenna for THz wireless communication," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 3, no. 5, pp. 711–716, 2014.

[15] O. Paul, C. Imhof, B. Reinhard, R. Zengerle, and R. Beigang, "Negative index bulk metamaterial at terahertz frequencies," *Optical Society of America*, vol. 16, no. 9, pp. 6736–6744, 2008.

[16] G. Kiziltas and J. L. Volakis, "Miniature antenna designs on metamaterial substrates," in *8th Int. Conference on Electromagnetics in Advanced Appl. (ICEAA), Conf. Proceedings*, pp. 431–434, Torino, Italy, 2003.

[17] N. Abdullah and G. Bhardwaj, "Design of squared shape SRR metamaterial by using rectangular microstrip patch antenna at 2.85 GHz," in *2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 196–200, Amity University, Noida, Delhi, February 2017.

[18] S. Kar, M. Ghosh, A. Kumar, and A. Majumder, "Complementary split ring resonator-loaded microstrip patch antenna useful for microwave communication," *World Academy of Science, Engineering and Technology, International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*, vol. 10, no. 10, pp. 1321–1325, 2016.

[19] L. Peng, P. Chen, A. Wu, and G. Wang, "Efficient radiation by electrically small antennas made of coupled split-ring resonators," *Scientific Reports*, vol. 6, no. 1, 2016.

[20] D. Chaturvedi and S. Raghavan, "SRR-loaded metamaterial-inspired electrically-small monopole antenna," *Progress In Electromagnetics Research C*, vol. 81, pp. 11–19, 2018.



Springer

KSME

Journal of Mechanical Science and Technology 33 (1) (2019) 1~10
www.springerlink.com/content/1738-494x(Print)/1976-3824(Online)
DOI 10.1007/s12206-018-1201-x

Vibration analysis for detecting failure mode and crack location in first stage gas turbine blade †

S. Rani*, A. K. Agrawal and V. Rastogi

Department of Mechanical, Production, Industrial and Automobile Engineering, Delhi Technological University, Delhi-110042, India

(Manuscript Received November 9, 2017; Revised May 23, 2018; Accepted August 20, 2018)

Abstract

Structure frequency response testing "modal analysis" is an integral part of the development and testing of structures such as pistons, turbine blades, compressor blades, crankshafts, and connecting rods. The usefulness of this technique lies in the fact that the energy in an impulse input is distributed continuously in the frequency domain. Thus, an impulse force will excite all resonances within given frequency range. To detect a fault in the structure, one may require frequency response functions (FRFs) of structures in both conditions, before (healthy structure) and after (failed structure) fault occurs. Now by extracting modal properties from collected FRFs and by comparing modal properties, one can detect and locate the structural faults. A case study is presented in order to detect failure mode and locate cracks on a 30 MW first stage gas turbine blade made of nickel based super alloy IN738LC, which has failed after rendering a useful life of 72000 h. The root causes of failure are detected by comparing the failed blade experimental model with the failed blade computational model. It is observed that the frequencies of the real failed blade experimental model are lesser than the computational model of the failed turbine blade. This is due to the metallurgical defects, which result in loosening of stiffness at the leading and trailing edges of the blade. Further, the stress concentration areas noticed on leading and trailing edges in computational model of the failed blade at the sixth mode are well corroborated with the cracked zone seen on leading and trailing edges of a real case failed turbine blade, collected from the site. It is concluded that the blade has failed due to that the resonance at sixth modal frequency. Scanning electron microscope (SEM) images reveal the presence of corrosion pits on the surfaces of the turbine blade that lead to surface degradation, which results in crack initiation and its propagation with high-cycle fatigue. It is concluded that the failure of turbine blade occurs due to high cycle fatigue.

Keywords: Vibration analysis; First stage IN738 gas turbine blade; Global rational fraction polynomial method (GRFP); Modal properties; Frequency response function (FRF)

1. Introduction

Modal parameters are global properties of a structure. Any changes in these parameters are used as a tool for damage detection. The structural faults affect mass, damping and stiffness properties of structures [1]. A number of methodologies have been proposed for damage detection in a structure based on the changes of modal parameters [2, 3]. Cawley et al. [4] used the ratio of frequency changes in two modes as a damage indicator. To locate the defect on the structure, the theoretical frequency shifts due to damage are calculated and compared with measured values. Pandey et al. [5] used the change in flexibility matrix due to the occurrence of damage in the structure as an indicator to locate the structural damage. Law et al. [6] used a finite-element model of the structure for determining changes in structural stiffness and positions of defects from measured changes in resonant frequencies. Shi et al. [7] located the structural damage by indicating the change in modal strain energy before and after the occurrence of damage in the structure. Khoshnoudian et al. [8] presented a global algorithm for damage assessment of structures, based on a parameter estimation method, using the finite element and measured modal response of the structure. Lin [9] multiplied the measured flexibility matrix by the undamaged analytical stiffness matrix. The resulting matrix should be an identity matrix if there is no damage. Salawu [10] proposed a global integrity index for detecting changes in the structural integrity of an engineering system. The index represents a linear relation among the frequencies of damaged structure and the undamaged structure. This index is capable of detecting changes in the global structural stiffness.

Researchers have performed beam vibration analysis to evaluate its modal properties (Boudjemai et al. [11], Yinming et al. [12], Khan et al. [13], Singh et al. [14, 15], Leticia et al. [16] and Elkhier et al. [17]).

*Corresponding author. Tel.: +98 735 75460
E-mail address: srani3508@gmail.com
† Recommended by Associate Editor Gyuhae Park
© KSME & Springer 2019

Various researchers evaluated modal parameters of turbine blade by experimental modal analysis and finite element analysis; changes found in modal parameters are used as a tool for finding damages and causes of failures. Choi et al. [18] carried out failure analysis of turbine blade. They evaluated modal parameters of failed blade experimentally and numerically. By comparing they have concluded that the fatigue fracture of the blade was originated during transient events internal to the combustion chamber. Rao et al. [19] performed vibration analysis of failed compressor blade and found that the blade failed due to high cycle fatigue mechanism. Poursaeidi et al. [20] presented a failure analysis of compressor blade experimentally and computationally. The computational results showed that high concentration of stresses occurred in the areas where the cracks existed in the real failed blade. They concluded that the first and second natural frequency modes are the primary reason for the fatigue failure of the blade. Kim et al. [21] prevented high cycle fatigue fracture of a 5 MW gas turbine by comparing the modal characteristic and resonance stability according to the tip shape and assembly condition. Madhavan et al. [22] detected the blade damage on the basis of abnormalities observed in the vibration characteristics of the blade tip timing data measured during engine tests.

In this paper a case study is presented to detect failure modes and crack locations in a failed 30 MW first stage gas turbine blade made up of nickel based super alloy IN738LC, which has failed after useful service of 72000 hours. Modal and harmonic analysis of the blade has been performed using *ANSYS*® software. Further turbine blade is analyzed with vibration analyzer OROS® to get frequency response functions (FRFs) of blade, which are obtained as output results by applying input artificially on the blade through impact hammer [23]. The modal properties of structures are extracted from resultant FRFs through global rational fraction polynomial method (GRFP) [24]. The root causes of failure of a gas turbine blade are detected by comparing the failed blade experimental model with the failed blade computational model. Turbine blade was also analyzed through scanning electron microscopy (SEM) to detect the cause of failure.

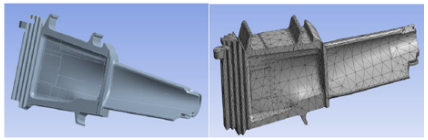
It is observed that the natural frequencies of a real case failed blade obtained experimentally are lesser than failed blade computational model. This is due to the metallurgical defects, which occur in real time failed blade as it failed after a service life of 72000 h and results in loosening its stiffness at the leading and trailing edges of the blade [25]. It is also analyzed that the stress concentration areas noticed on leading and trailing edges in computational model of the failed blade at the 6th mode are well corroborated with the cracked zone seen on leading and trailing edges of a real case failed turbine blade, which provides a basis to conclude that the blade has failed due to resonance at sixth modal frequency. Scanning electron microscopy (SEM) images reveal the presence of corrosion pits on the surfaces of turbine blade [26] that leads to surface degradation, which results in crack initiation and its propagation with high-cycle fatigue. It is concluded that the failure of blade takes place due to high cycle fatigue mechanism.

Table 1. Material properties of turbine blade and fixture (source- technical data- NiPERA) [35].

Material properties	Turbine blade
Young's modulus (N/m^2)	1.3996x10^{11}
Poisson's ratio	0.3
Density (Kg/m^3)	8110
Bulk modulus (N/m^2)	1.1664x10^{11}
Shear modulus (N/m^2)	5.3832x10^{10}
Tensile yield strength (N/m^2)	3.4474x10^8
Tensile ultimate strength (N/m^2)	4.5505x10^8

Table 2. Dimensions of turbine blade and fixture.

Dimensions	Turbine blade (size of bounding box)
Length, L (m)	0.25
Breadth, b (m)	0.11
Depth, h (m)	0.054



(a) (b)

Fig. 1. (a) 3-D Model of turbine blade; (b) mesh model of turbine blade.

The paper is organized in five sections. Sec. 2 represents the computational analysis of a real case failed turbine blade. Sec. 3 provides the details of experimental set-up used for experimental modal analysis of blade. Sec. 4 provides method of evaluation of modal properties of blade through GRFP method. Sec. 5 includes results and discussion. Sec. 6 contains the conclusions.

2. Computational modal and harmonic analysis of turbine blade

In this paper, modal and harmonic analysis of a real case failed turbine blade was performed using *ANSYS*® software. A three-dimensional CAD model was generated through a 3D scanning process so that the dimensions, contour and blade profiles are accurately and efficiently captured. The 3D data set of the scanned turbine blade was converted to a solid model as shown in Fig. 1(a). The triangular surface mesh was made followed by volumetric meshing using auto meshing features of *ANSYS® 15* as shown in Fig. 1(b). The finite element model generated has 42538 nodes and 25328 elements. The boundary condition was provided by applying fixed support at fir-tree part of turbine blade. The material properties and dimensions of turbine blade are listed in Tables 1 and 2, respectively.

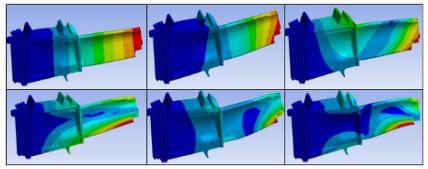


Fig. 2. Six modes of failed gas turbine blade.

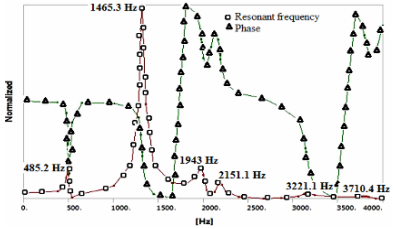


Fig. 3. Frequency response function (FRF), displacement vs frequency.

2.1 Modal analysis of the failed gas turbine blade

Computational modal analysis of the failed turbine blade was performed on *ANSYS*® software; the collected values of first six resonant frequencies of failed blade were evaluated as *485.2 Hz*, *1465.3 Hz*, *1943 Hz*, *2151.1 Hz*, *3221.1 Hz* and *3710.4 Hz*, their respective mode shapes are shown in Fig. 2.

2.2 Harmonic analysis of the failed gas turbine blade

Frequency response function (FRF) as shown in Fig. 3 was obtained by performing harmonic response analysis of the failed turbine blade on *ANSYS*® software. "Peak" responses are identified on the plots. These peak values are corresponding to natural frequencies of the turbine blade. The amplitude vs frequency FRF is shown by square line. The first six resonant frequencies are at *485.2 Hz*, *1465.3 Hz*, *1943 Hz*, *2151.1 Hz*, *3221.1 Hz* and *3710.4 Hz*. The phase angle vs frequency FRF is shown by triangle line. It is clearly evident from Fig. 3 that the response is changing its sign at resonant frequencies.

The resonant frequencies and mode shapes of the failed turbine blade obtained computationally are validated experimentally in the next section to find the root cause of turbine blade failure.

3. Experimental modal analysis of a failed turbine blade

The experimental modal analysis was carried out on a real case failed turbine blade, collected from the gas power plant site, by using vibration analyzer *OROS*® with impact testing**.** The detailed steps involved in modal analysis of turbine blade are represented in Fig. 4.

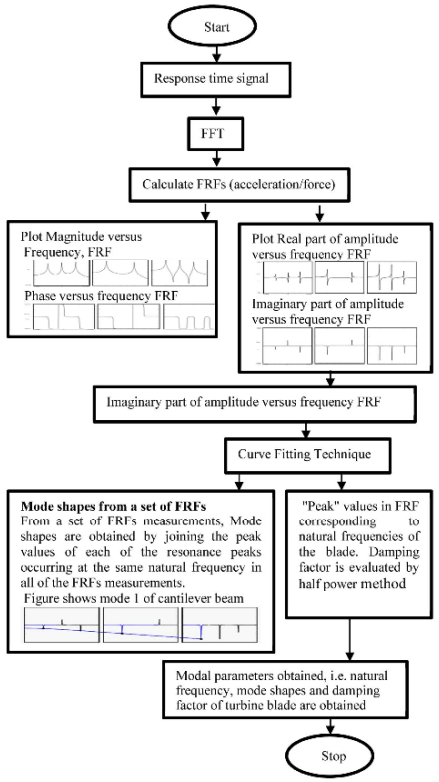


Fig. 4. Methodology for experimental modal analysis of turbine blade.

3.1 Fabrication of turbine blade fixture

Frequency response functions (FRFs) of the turbine blade are measured by making one end of the turbine blade fully built-in; the boundary conditions are similar to that of a cantilever. This boundary condition can be established by using a fixture. The turbine blade fixture was fabricated on a universal milling machine with a bevel protector at an angle of 11.5^o and an up-milling cutter mounted in the arbor with wedge face perpendicular to it, which is shown in Fig. 5. The turbine blade fixture consists of three separate parts, including two mild steel blocks and a hexagonal bolt and nut. One side both the mild steel blocks has fir-tree profile cut on universal mill-



Fig. 5. Universal milling machine.

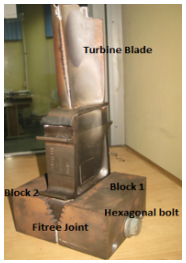


Fig. 6. The turbine blade with fixture.

ing machine and both the blocks have through holes at the bottom of fir-tree profile. Both blocks are placed side by side and after placing the turbine blade, a hexagonal bolt is used to lock the whole assembly with the help of a hexagonal nut as shown in Fig. 6.

3.2 Measurement setup

The experimental modal analysis setup, shown in Fig. 7, consists of the turbine blade, a turbine blade fixture, a *PCB-78534* accelerometer, a *PCB-086C03* impact hammer and a vibration analyzer *OROS*®. First, the hammer tip (plastic), the inputs (force and acceleration), the frequency range (*0-5 KHz*), triggering (start delay: *-10 ms*), windows (force: hamming and acceleration: response), sampling rate (*51.2 kilo samples / sec*) and the number of FFT lines *(401)* are selected. The accelerometer is placed at the tip of trailing edge.

3.3 Measurements points

For modal analysis of turbine blade, 26 measurement points were taken, distributed across the convex side of the turbine blade as shown in Fig. 8. Along the length of the turbine blade, six measurement location points are marked from 1 to 6. These location points are equally spaced at a distance of 4.4 cm from top to bottom. Across the width of the blade, six measurement points were taken at locations 1, 2 and 3. For locations 4 to 6, three measurement points were taken because

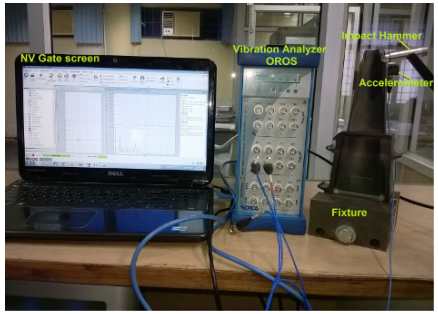


Fig. 7. Experimental modal analysis of failed gas turbine blade.



Fig. 8. Measurement points on turbine blade.

the blade is tapered at the top. The distance between these horizontal measurements points varies at each location. The accelerometer is placed at location point 1 on the tip of trailing edge and the impact force is applied at all measured locations marked on the blade to define mode shapes of the turbine blade.

3.4 Recorded signal

The input signal is recorded for 10 seconds. Recorded signals were obtained by hitting impact hammer at measured locations. The trigger is set at start delay of *-10 ms*. Recorded signal and triggered block of signal at measurement location 1 is shown in Figs. 9(a) and (b), respectively.

3.5 Frequency response functions (FRFs)

The frequency response function (FRF) plot is the representation of the ratio of output to the input signal Vs frequency. Fig. 10 presents the FRF plot of (Acceleration/Force) Vs frequency measured at location 1. It is clearly evident from Fig. 10 that the response is changing its sign at resonant frequencies. The frequency response functions (FRFs) plots are further used to calculate modal damping and mode shape of blade.

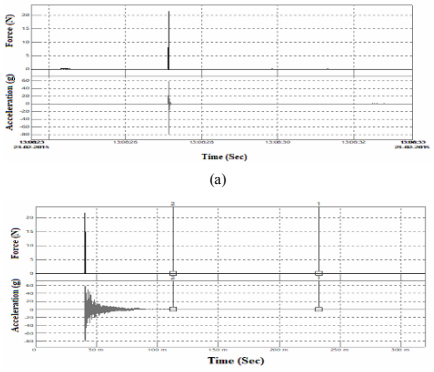


Fig. 9. (a) Recorded signal measured at location 1: (b) triggered signal measured at location 1.

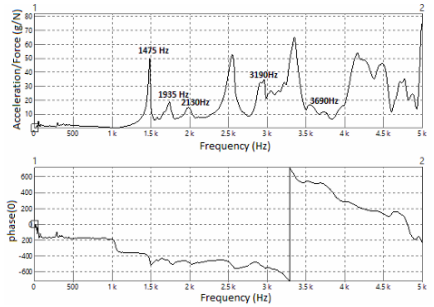


Fig. 10. FRF measured at location 1.

4. Estimation of modal parameters of turbine blade

The modal parameters, i.e., natural frequency, damping and mode shapes of turbine blade, are evaluated from FRFs by using global rational fraction polynomial method (GRFP) method. The next subsections present the extraction of turbine blade model parameters from collected frequency response functions (FRFs).

4.1 Modal frequency

"Peak" values seen in FRF (Fig. 10) are corresponding to natural frequencies of the blade. The six dominant resonance peaks are clearly shown in FRF at frequencies around *484 Hz, 1475 Hz, 1935 Hz, 2130 Hz, 3190 Hz* and *3690 Hz*. Therefore, the first six natural/ resonant frequencies of turbine blade are *484 Hz, 1475 Hz, 1935 Hz, 2130 Hz, 3190 Hz* and *3690 Hz*.

It is observed in experimentally obtained FRF's that the peaks are also seen at around *50 Hz, 300 Hz, 1650 Hz, 2550 Hz, 2950 Hz* and *3350* Hz. However, these frequencies cannot be identified as resonance peaks because these frequencies are not seen in computationally obtained frequency response functions (FRFs). These peaks may be observed due to surface defects found on the investigated failed turbine blade [26].

Table 3. Modal damping coefficients of turbine blade.

Mode	ω_1 (Hz)	ω_2 (Hz)	ω_k (Hz)	ζ
1	494	478	484	0.0165
2	1500	1450	1475	0.0169
3	1962	1898	1935	0.0165
4	2160	2089	2130	0.0166
5	3243	3135	3190	0.0169
6	3747	3625	3690	0.0165

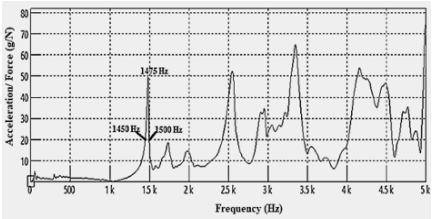


Fig. 11. Modal damping co-efficient evaluated at natural frequency *1475 Hz* using half power method.

4.2 Modal damping

Modal damping is evidenced by the width of the resonance peaks. Fig. 11 represents the damping co-efficient evaluated at 2nd modal frequency *1475 Hz* by applying half power method. The calculated values of all six damping coefficients as per the method represented in Fig. 11 are tabulated in Table 3.

4.3 Mode shape

The plots of the real and imaginary magnitudes of response versus frequency are most useful when dealing with modes. The resonance occurs as peaks in the imaginary magnitude plot when the real magnitude passes through zero. The real and imaginary plot of FRF at measured location 1 is shown in Fig. 12.

To obtain the mode shapes of turbine blade, consider the imaginary part of all the FRFs obtained from location 1 to location 6 as shown in Fig. 13. The mode shapes of the blade are obtained by joining the peak of each of the resonance

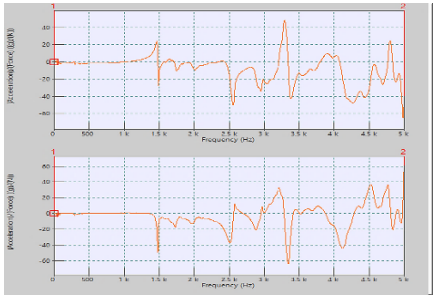
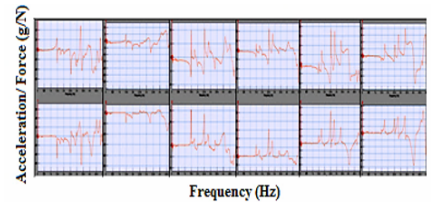


Fig. 12. Real and imaginary plots of FRF measured at location 1.



Location 1 Location 2 Location 3 Location 4 Location 5 Location 6

Fig. 13. Real (top) and Imaginary (below) plots of collected FRF's from measured locations 1 to 6.

peaks which occur at the same frequency in all the measured FRFs from location 6 to location 1.

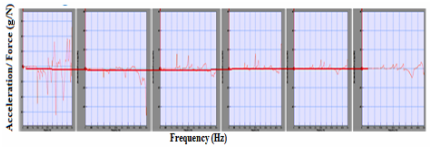
Mode 1 of the turbine blade is obtained by joining peak values at natural frequency *484* Hz at all the measured locations starting from the measured location 6 to the measured location 1. Similarly, mode shapes 2, 3, 4, 5 and 6 are obtained by joining peak values at natural frequency *1475 Hz, 1935 Hz, 2130 Hz, 3190 Hz* and *3690 Hz* at all the measured locations starting from the measured location 6 to the measured location 1 respectively. All six mode shapes are shown in Figs. 14(a)-(f).

5. Results and discussion

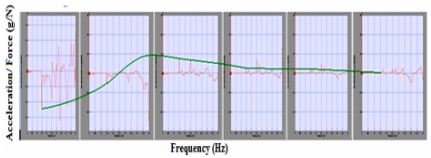
The root causes of failure of the blade are detected by comparing the modal analysis results of the failed turbine blade obtained experimentally and computationally.

While comparing the experimental model of failed blade with the computational model of failed blade, it is observed that the natural frequencies of the real case failed blade obtained experimentally are approximately the same as per computational model of failed blade. The marginal difference is due to the metallurgical defects, which occur in the real time failed blade as it fails after a service life 72000 hrs and results in loosening of its stiffness at the leading and trailing edges.

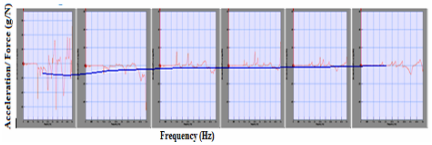
Since the investigated blade (failed blade) is a first stage high pressure blade, this is under the influence of high temperature and pressure. These ambient conditions result in changing of microstructure of blade material. The investigated



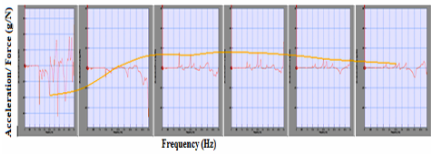
(a) Mode 1 at frequency 484 Hz



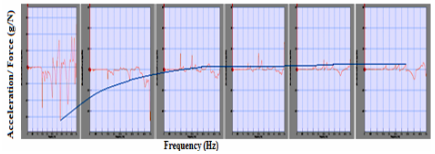
(b) Mode 2 at frequency 1475 Hz



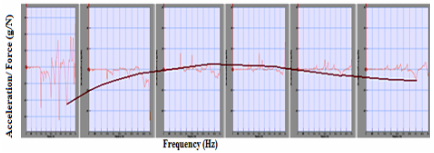
(c) Mode 3 at frequency 1935 Hz



(d) Mode 4 at frequency 2130 Hz



(e) Mode 5 at frequency 3190 Hz.



(f) Mode 6 at frequency 3690 Hz

Fig. 14. Mode shapes of gas turbine blade obtained by applying GRFP method.

Table 4. Comparison of natural frequencies of computational failed blade model and experimental of failed blade model.

Mode	Harmonic modal analysis (HMA) of failed turbine blade, frequency [Hz]	Experimental modal analysis (EMA) of failed turbine blade, frequency [Hz]	Error [%] b/w HMA results & EMA results
Mode 1	485.2	484	0.0024
Mode 2	1465.3	1475	-0.0068
Mode 3	1943	1935	0.0041
Mode 4	2151.1	2130	0.0098
Mode 5	3221.1	3190	0.0096
Mode 6	3710.4	3690	0.0054

blade was manufactured with nickel based superalloy IN738. Ni based super-alloys have two-phase equilibrium microstructures, consisting of gamma (γ) and gamma prime (γ') phases [27, 28]. Both the phases have a face-centered cubic (FCC) lattice with similar lattice parameters [29]. Since their lattice parameters are similar, the γ' is coherent with the γ when the size of precipitate is small. Under high temperature deformation, the finely dispersed γ' phase changes to a layered structure, which is perpendicular to the applied uniaxial stress; this results in breaking of the overall cubic symmetry of the precipitate shape. Crack growth suddenly occurs in the grain along the interface between the layered γ' phase and γ phase, and results in the reduction of stiffness of blade [30].

The calculated percentage difference of natural frequencies between failed blade experimental and computational models becomes negative for the second mode because the experimentally determined shape of the second mode came from the lateral motion of the fixed blade and the fixture could not fully constrain the lateral motion [18].

The 3-D scanned computational model of failed blade has deviation in geometry due to cracks found on the leading and trailing edges of collected turbine blade, but no metallurgical defects are considered in computational model.

The mode shapes of the failed blade experimental model are well corroborated with the turbine blade computational model as shown in Figs. 15-20.

It is also found that the stress concentration areas noticed on leading and trailing edges in computational model of the failed blade at the 6th mode are well corroborated with the cracked zone seen on leading and trailing edges of a real case failed turbine blade as shown in Figs. 21 and 22, which provides a basis to conclude that the blade has failed due to resonance at sixth modal frequency.

Generally, modal analysis is carried out before the structure is placed in field applications. By evaluating the modal parameters of turbine blade, resonance can be avoided or can be brought to a lesser percentage. Failure percentage of blade goes down due to reduction in resonant vibration. On the basis of modal analysis results, one may likely take remedial actions prior to failure.

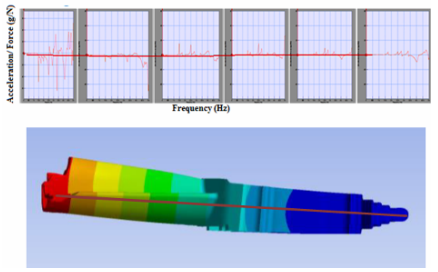


Fig. 15. Experimental and computational obtained mode 1.

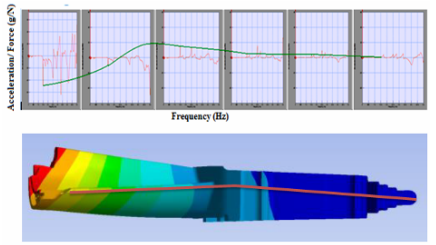


Fig. 16. Experimental and computational obtained mode 2.

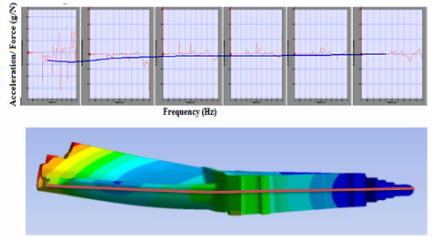


Fig. 17. Experimental and computational obtained mode 3.

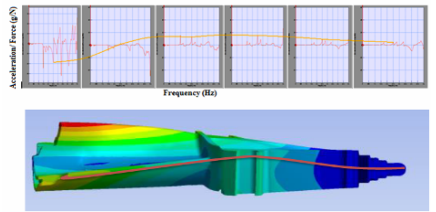


Fig. 18. Experimental and computational obtained mode 4.

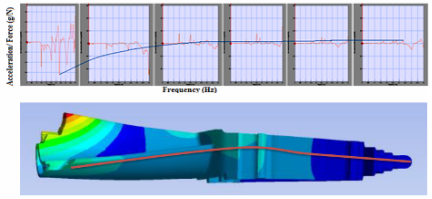


Fig. 19. Experimental and computational obtained mode 5.

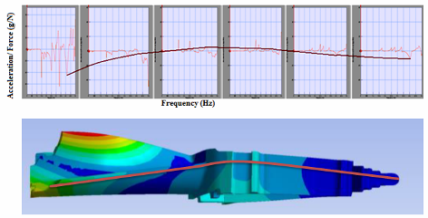


Fig. 20. Experimental and computational obtained mode 6.

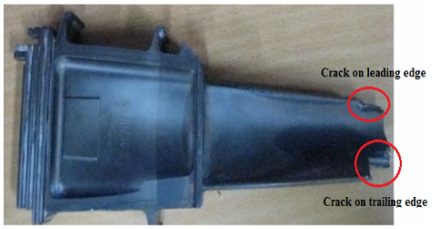


Fig. 21. Failed first stage high pressure gas turbine blade collected from site.

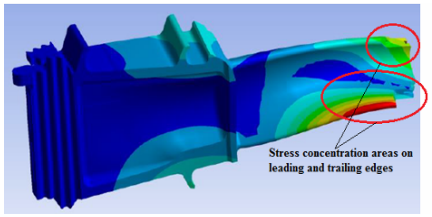


Fig. 22. Mode 6 of failed blade (bending and twisting) at frequency 3710.4 Hz.

Scanning electron microscopy (SEM) images reveal the presence of pits on the surfaces of turbine blade as observed in Fig. 23. The corrosion products penetrate deep into metal through these initiated cracks, which results in crack propagation. During crack propagation, hot corrosion occurs [31, 32] and causes enlargement of the crack area. At high temperature,

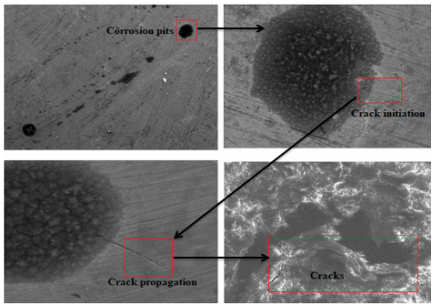


Fig. 23. SEM images of corrosion pits present over the trailing edge.

rapid oxidation takes place in the presence of sodium sulfate, which combines with chromium to form a mixture of salts and sulfur. This released sulfur penetrates inwards and reacts with chromium to form chromium sulfide in the substrate. Further, sulfides are converted to complex unstable metal oxides as corrosion proceeds. This sulfidising of the substrate forms the nickel sulfide eutectic at the grain boundaries, having lower melting point than the working temperature of the blade. This leads to incipient melting of the grain boundaries and lowers the strength of the blade drastically. Once the corrosive compounds have eaten away the grain boundaries, cracking of the blade gets started by high cycle fatigue [33, 34].

6. Conclusions

Vibration analysis of turbine blade was carried out for detecting failure mode and crack location. The following conclusions were drawn:

(i) The novelty of this approach lies in the fact that the experimental analysis has been carried out on a real case failed turbine blade and the computational analysis has been considered for the three-dimensional CAD model generated through a 3D scanning process of the same failed turbine blade so that dimensions, contour and blade profiles are accurately and efficiently captured.

(ii) The root causes of failure of a gas turbine blade are detected by comparing the failed blade experimental model with the failed blade computational model. By comparing the failed blade experimental model with the failed blade computational model, it is observed that the natural frequencies of a real case failed blade obtained experimentally are lesser than failed blade computational model. This is due to the metallurgical defects, which occur in real time failed blade as it failed after a service life of 72000 h and results in loosening its stiffness at the leading and trailing edges of the blade.

(iii) It is also analyzed that the stress concentration areas noticed on leading and trailing edges in computational model of failed blade at the 6th mode are well corroborated with the cracked zone seen on leading and trailing edges of a real case failed turbine blade, which provide a basis to conclude that the blade has been failed due to resonance at sixth modal frequency.

(iv) Scanning electron microscopy (SEM) images reveal the presence of corrosion pit on the surfaces of turbine blade that leads to surface degradation, which results in crack initiation and its propagation with high-cycle fatigue. It is concluded that failure of turbine blade occurred due to high cycle fatigue.

(v) To avoid fatigue failure of turbine blade, the use of under platform friction damper is being proposed for the fatigue life enhancement of turbine blade. Friction damper dissipates energy in the form of heat due to the rubbing motion of the contacting surfaces resulting from the relative motion and thus reduce vibratory stresses [35].

References

[1] M. A. Mannan and M. H. Richardson, Detection and location of structural cracks using FRF measurements, *IMAC VIII* (1990) 1-6.

[2] T. Nestrorovic, M. Trajkov and M. Patalong, Identification of modal parameters for complex structures by experimental modal analysis approach, *Advances in Mechanical Engineering,* 8 (2016) 1-16.

[3] O. S. Salawu, Detection of structural damage through changes in frequency: A review, *Engineering Structures*, 19 (1997) 718-723.

[4] P. Cawley and R. D. Adams, the location of defects in structures from measurement of natural frequencies, *J. of Strain Analysis,* 14 (1979) 49-57.

[5] A. K. Pandey and M. Biswas, Experimental verification flexibility difference method for locating damage in structures, *J. of Sound and Vibrations*, 184 (1995) 311-328.

[6] S. S. Law, L. Xun and H. S. Ward, A vibration technique for structural stiffness identification, *Proceedings of International Conference on Vibration Problems in Engineering* (1990) 683-698.

[7] Z. Y. Shi, S. S. Law and L. M. Zhang, Structural damage detection from modal strain energy change, *J. of Engineering Mechanics*, 126 (2000).

[8] F. Khoshnoudian and A. Estandiari, Structural damage diagnosis using modal data, *Scientia Iranica*, 18 (2011) 853-860.

[9] C. S. Lin, Location of modeling errors using modal test data, *AIAA J.*, 28 (1990) 1650-1654.

[10] O. S. Salawu, An integrated index method for structural assessment of engineering structures using modal testing, *J. of the British Institute of Non Destructive Testing*, 39 (1997).

[11] A. Boudjemai, R. Amri and A. Mankour, Modal analysis and testing of hexagonal honeycomb plates used for satellite structural design, *J. of Materials and Design,* 35 (2012) 266-275.

[12] S. Yinming, H. Hongxing and S. Hugo, The finite element analysis and experimental study of beams with active constrained layer damping treatments, *J. of Sound and Vibrations,* 278 (2004) 343-363.

[13] I. A. Khan and D. R. Parhi, Finite element analysis of double cracked beam and its experimental validation, *Procedia Engineering,* 53 (2013) 703-708.

[14] B. Singh and B. K. Nanda, Dynamic analysis of damping in layered and welded beams, *J. of Engineering Structures*, 48 (2013) 10-20.

[15] B. Singh and B. K. Nanda, Dynamic analysis of damping mechanism in welded multilayered mild steel beams, *J. of Aerospace Science and Technology,* 29 (2013) 58-73.

[16] F. Letícia and M. Fadel, Theoretical and experimental modal analysis of a cantilever steel beam with a tip mass, *J. of Mechanical Engineering Science*, 223 (2009) 1535-1541.

[17] M. A. A. Elkhier-Abo, A. Hamada and A. Bahei El-Deen, Prediction of fatigue life of glass fiber reinforced polyester composites using modal analysis, *International J. of Fatigue*, 69 (2014) 28-35.

[18] Y.-S. Choi, Investigation of blade failure in a gas turbine, *J. of Mechanical Science and Technology*, 24 (2010) 1969-1974.

[19] A. R. Rao and B. K. Dutta, Vibration analysis for detecting failure of compressor blade, *J. of Engineering Failure Analysis*, 25 (2012) 211-218.

[20] E. Poursaeidi, A. Babaei and M. Arablu, Effects of natural frequencies on the failure of R1 compressor blades, *J. of Engineering Failure Analysis*, 25 (2012) 304-315.

[21] K. K. Kim and Y. S. Lee, Modal characteristics according to the tip shape and assembly condition of turbine blade, *J. of Mechanical Science and Technology*, 27 (2013) 3409-3417.

[22] S. Madhavan, R. Jain, C. Sujatha and A. S. Shekar, Vibration based damage detection of rotor blades in gas turbine engine, *J. of Engineering Failure Analysis*, 46 (2014) 26-39.

[23] M. H. Richardson and B. J. Schwarz, *Experimental modal analysis*, CSI Reliability Week (1999) 1-12.

[24] M. H. Richardson and D. L. Formenti, Global curve fitting of frequency response measurements using the rational fraction polynomial method, *Proceedings of the 3rd International Modal Analysis Conference Orlando,* Florida (1985) 390-397.

[25] M. H. Richardson and M. A. Mannan, *Correlating minute structural faults with changes in modal parameters*, IMAC XI (1993) 1-7.

[26] S. Rani, A. K. Agrawal and V. Rastogi, Failure analysis of a first stage IN738 gas turbine blade tip cracking in a thermal power plant, *Case studies of Engineering Failure Analysis,* 8 (2017) 1-10.

[27] R. J. Mitchell and M. Preuss, Inter-relationships between composition, γ' morphology, hardness, and γ-γ' mismatch in advanced polycrystalline nickel-base super alloys during aging at 800 °C, *Metallurgical and Materials Transactions A,* 38 (2007) 615-627.

[28] G. X. Yang, Y. F. Xu, L. Jiang and S. H. Liang, High temperature tensile properties and fracture behavior of cast nickel-base K445 super alloy, *Progress in Natural Science*, 21 (2011) 418-425.

[29] J. Li and R. P. Wahi, Investigation of γ/ γ' lattice mismatch

in the polycrystalline nickel-base superalloy IN738LC: Influence of heat treatment and creep deformation, *Acta Metallurgica et Materialia*, 43 (1995) 507-517.

[30] J. Zrnik, P. Strunz, M. Maldini, A. Wiedenmann and V. Davydov, Small-angle neutron scattering investigation of γ' precipitate morphology evolution in creep-exposed single-crystal Ni-base superalloy CMSX-4, *J. of Physics: Condensed Matter*, 20 (2008) 104261-1-104261-8.

[31] N. Eliaz, G. Shemesh and R. M. Latanision, Hot corrosion in gas turbine component, *Engineering Failure Analysis,* 9 (2002) 31-43.

[32] M. T. Naeem, S. A. Jazayeri and N. Rezamahdi, Failure analysis of gas turbine blades, *Proceedings of the* 2008 *IAJC-IJME International Conference 120 ENG 108* (2008).

[33] R. K. Mishra, K. S. Jayasimha, J. K. K. George and A. Iftekar, Fretting failure of turbine blade in turbojet engine: A case study, *Engineering Note, Paper code V63* N3/734-2011 (2011).

[34] S. Rani, A. K. Agrawal and V. Rastogi, Analysis of first stage gas turbine blade for evaluating damping energy and failure mode, *Indian Journal of Science and Technology,* 10 (2017) 1-6.

[35] L. K. Bhagi, V. Rastogi and P. Gupta, Study of corrosive fatigue and life enhancement of low pressure steam turbine blade using friction damper, *J. of Mechanical Science and Technology*, 31 (2017) 17-27.



S. Rani has her Ph.D. in Machine Design (Vibrations) from Delhi Technological University, Delhi, India. Presently, she is working as Assistant Professor in Mechanical, Production, Industrial and Automobile Engineering Department at Delhi Technological University, Delhi, India. Her research interest is fault diagnosis in Machinery.

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Received: 10 October 2018
Accepted: 13 December 2018
Published online: 09 January 2019

Wavefunction Engineering of Type-I/Type-II Excitons of CdSe/CdS Core-Shell Quantum Dots

Yashaswi Nandan & Mohan Singh Mehata

Nanostructured semiconductors have the unique shape/size-dependent band gap tunability, which has various applications. The quantum confinement effect allows controlling the spatial distribution of the charge carriers in the core-shell quantum dots (QDs). Upon increasing shell thickness (e.g., from 0.25–3.25 nm) of core-shell QDs, the radial distribution function (RDF) of hole shifts towards the shell suggesting the confinement region switched from Type-I to Type-II excitons. As a result, there is a jump in the transition energy towards the higher side (blue shift). However, an intermediate state appeared as pseudo Type II excitons, in which holes are co-localized in the shell as well core whereas electrons are confined in core only, resulting in a dual absorption band (excitation energy), carried out by the analysis of the overlap percentage using the Hartree-Fock method. The findings are a close approximation to the experimental evidences. Thus, the understanding of the motion of *e-h* in core-shell QDs is essential for photovoltaic, LEDs, *etc*.

Quantum dots primarily exhibits the confinement effect that leads to the spatial enclosure of the electronic charge carriers within the nanocrystals (NCs)[1]. The effect is present when the dimensions of the semiconductor are roughly below the Bohr's exciton radius of that specific enclosure. This quantizes the conduction band (CB) and the valence band (VB) energy levels, nearly continuous for bulk semiconductors. The spacing between the valence band and conduction band can be varied by changing the size or shape of the nanocrystals which gives more monochromatic emission and near unity quantum yield by increasing shell thickness. At strong confinement regime (size below the Bohr's exciton radius) allows for controlling the spatial distribution of carrier wave functions across various domains of a hetero-nanostructure QDs which shows size-dependent absorption and photoluminescence (PL) spectra[2,3].Within a study carried out, the calculated exciton's binding energy approximates to the reported Bohr's exciton radius of CdS single crystal (2.5 nm)[4] at the shell thickness of 2.4 nm, as obtained in our model.

The spatial separation of the electron and hole wave functions within such semiconductor heterostructure[5,6] results in a prolonged charge transfer (CT) states which favours desirable characteristics for applications, such as light emitters, lasers, photo-catalysts and photovoltaic devices[7–12]. The charge separation state can be achieved either by direct excitation of the CT state or by photoexcitation of an exciton followed by hole transfer through the core-shell boundary. Investigating the formation and relaxation of the CT state provides information about the excitation and de-excitation processes[13]. Recently, the hetero-structuring idea has been implemented by various groups for actuating NCs capability of emitting multi-colour light because of radiative recombination of excitons confined into core/shell domains. These are DiB's (Dot in bulk) semiconductors, which is due to the large size of the shell as compared to the core exhibits dual emission. The primary reason for dual colour emission is believed to be the size of the shell, the less non-radiative transfer of the shell's hole towards the core, enhancing the probability of radiative recombination of the shell's exciton into dual emission[14]. Also, the geometric separation of excitons in the nanostructures results in inhibited relaxation processes (exciton cooling and multi-exciton recombination), which gives dual-emission[15].

Materials which has small lattice mismatch, the strain-induced localization of carriers are less efficient, as the energy levels in the CB and VB shifts insignificantly[16]. This is the case with CdSe/CdS NCs, which are structures having specific properties such as mono-dispersity, narrow emission bandwidth, and high quantum yield[17,18]. In shape altered QDs (nano-rods), it has been reported that the smaller CB and VB offset of CdSe favour localization

Laser-Spectroscopy Laboratory, Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi, 110042, India. Correspondence and requests for materials should be addressed to M.S.M. (email: msmehata@gmail.com)

of both electron and hole inside the core (Type-I) exciton. However, as the core size decreases, the kinetic energy of electron permits to penetrate CB offset barrier, confining over shell (Type-II) exciton[13,19–21]. This effect is predominant on the graded interface which has a step-like confinement potential at the interface. However, this separation between electron and hole is usually restricted to small-sized core quantum mechanical systems and is limited to Coulomb interaction[1,13,22]. A previous study on the differential localization of the electron and hole in the core-shell QDs is limited in the insights of the overlap integral of charge carriers for the determination of the different types of excitons and its effects on the size of the QDs. For a better understanding of the dual excitations, bulk shell structures are also needed to be studied, as it follows the same trend as predicted by our proposed model on the limitations of the dual excitations for the shell sizes[23]. In another report[24] pseudopotential method is used to study the effect of the environment on the energy values of the electron. The environment induces excess charge carrier to imitate dielectric effect, which has been simplified by using Gauss's law on the Hamiltonian equation to extract eigenvalues accordingly.

Controlling the spatial distribution of the charge carrier in nanorods (NRs) by altering the environment and applying an electric field, keeping the dielectric of the rod unaltered is also reported[25]. Low insulating strengths of the environment results in increased *e-h* interactions thereby increasing binding energy along with the transition probability of Type-I NRs. For the Type-II NRs, the spatial distribution of the electron completely shifts towards shell, hence transitions from pseudo-Type-II to Type-II exciton.

A novel method has been devised[26] to compare tuning of the localization of the carrier by reduction using different photochemical *hole* quenchers (ethanol *vs*. borohydrides) of ZnO QDs. This facilitates the accumulation of *electron* in the conduction band at a rate faster than the Auger recombination without the change in charge carrier at equilibrium. By photochemical reduction using *hole* quenchers, it induces delayed response of the *e-h* recombination thereby increasing the quantum yield efficiency and the absorptivity of the QDs. A similar study[27] has been carried out for further investigation to differentiate the regimes of electronic doping, which produces excess charge carriers and the electrostatic redox effect with equilibrium excess charge carriers. Reduced QDs produced delocalized electrons, and these electrons are produced primarily due to thermal excitations as the Fermi level is very close to the conduction band. The other regime has a Fermi level above CB, which results in a significant change of the charge carriers.

Despite previous works[15,28] on the switch of the localization of the electron and hole in core-shell CdSe/CdS structure, these studies have ignored the explanation of the mechanism in the intermediate of e-h pair transition from Type-I exciton to Type-II exciton, specifically pseudo Type-II exciton.

Here, we have modelled 1-D core-shell (CdSe/CdS) quantum dots of radius 2–5 nm, with Hartree-Fock mean field approximation theorem solved using FDM (Finite Difference Method). The shortcoming of the Hartree-Fock Effective Mass Approximation (EMA) model in the Poisson equation was taken care of by calculating the specific e-*h* coulomb interactions without averaging it out across the domain. The mean field approximation method is limited to small QD systems. Therefore we adopted a method for calculation of Coulomb interactions known as Exact Field Approximation (EFA), which is not exploited sofar[23,28,29]. Further, the calculation of the excitation energy by the analysis of the PDF (Probability Density Function) of the electron and hole is used to discriminate the localization of the charge carriers in the core-shell domain for the size dependence variation. We studied the Schrodinger equation for the exciton, which is bounded by an electron and a hole and examined the effect of the shell thickness variation on the radial distribution functions (RDF) of both within the strong confinement regime in which the perturbation is calculated as the *e-h* Coulomb interactions. Our study focuses on the coverage of *e-h* couple from core to shell CdSe/CdS NCs that produces transitions from a Type-I to a quasi-Type-II regime and continues into the onset of the Type-II regime depending on the increase in the size of QDs.

The Model and Theory

The core material, taken as CdSe with the radius $\boldsymbol{R_1}$, which is passivated with a CdS shell of radius $\boldsymbol{R_2}$, has a broader band gap than CdSe. The shell thickness is $T_s = R_2 - R_1$. The conduction and valence band profile of this structure is shown in the Fig. 1. The EMA (effective mass approximation) calculations (Supporting information (SI) of Method and Calculations) and BenDaniel-Duke boundary conditions were applied to solve the partial differential equations, we have assumed a 1-D spherical DLQD (Dual-layer quantum dots) heterostructure.

Schrodinger equation of an exciton is given as follows[30]:

$$\left[-\frac{\hbar^2}{2}\nabla_e\left(\frac{1}{m_e^*(r)}\right)\nabla_e - \frac{\hbar^2}{2}\nabla_h\left(\frac{1}{m_h^*(r)}\right)\nabla_h - \frac{e^2}{\kappa|r_e - r_h|} + V_h(r) + V_e(r)\right]\psi_{nlm}^{exc}(r_e,\ r_h) = \varepsilon_{nlm}\psi_{nlm}^{exc}(r_e,\ r_h) \quad (1)$$

Here, the first and second terms are the carrier charge kinetic energies of the electron and hole, respectively. The subsequent term is the attractive Coulombic interaction energy between the electron and hole, ħ is the reduced Planck constant, $\boldsymbol{m_e^*(r)}$ and $\boldsymbol{m_h^*(r)}$ are the position dependent electron and hole effective masses, respectively. The effective masses of the electron and hole of the CdSe and CdS have calculated the density of state effective mass (see SI of Table S1). $\boldsymbol{e}$ is the unit coulomb charge, $\boldsymbol{\kappa}$ is the dielectric charge of the medium, and $\boldsymbol{V_e(r_e)}$ and $\boldsymbol{V_h(r_h)}$ are the confinement potentials of the electron and hole, respectively (SI of Table S1). The $\boldsymbol{\psi_{nlm}^{exc}}(r_e,\ r_h)$ are the excitonic wave functions and the ε_{nlm} are the exciton energies. The exact solution of this PDE (partial differential equation) require huge computing resources, as the number of interacting matrix elements is approximately in the order of $10^{52} \times 10^{52}$. This requires large storage setup, faster processing for large matrix calculation and diagonalization assuming terabytes of data flow. Therefore, some approximations have to be made[31]. In the said structure, the electron moves in a mean potential induced by the hole and similarly, the hole moves in a mean potential

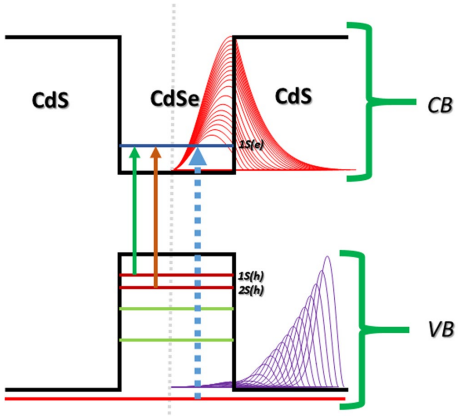


Figure 1. Band structure of core-shell CdSe/CdS QDs with the dual absorption (excitation) mechanism. The vertical solid line shows transitions between the core HOMO to LUMO and the dotted line between shell HOMO to core LUMO. The CB denotes the conduction band and VB denotes the valence band. Inset: PDF of electron (red) and hole (purple) are added for comparison w.r.t band structure. Demarcation dotted line in grey to point the core's centre.

induced by the electron[32]. In HF (Hartree-Fock) calculations, the carriers Schrodinger equations can be simplified as[30]:

$$\left[-\frac{\hbar^2}{2}\nabla_e\left(\frac{1}{m_e^*(r)}\right)\nabla_e - q_e\phi_h + V_e(r)\right]R_e(r) = \varepsilon_e R_e \quad (2)$$

$$\left[-\frac{\hbar^2}{2}\nabla_h\left(\frac{1}{m_h^*(r)}\right)\nabla_h - q_h\phi_e + V_h(r)\right]R_h(r) = \varepsilon_h R_h \quad (3)$$

In these equations, q_e and q_h are the charges, and $R_e(r)$ and $R_h(r)$ represents the radial wavefunctions of the electron and hole, respectively. $\varphi_e(\varphi_h)$ is the Coulombic potential generated by the electron (hole). The two PDEs are used to solve the radial part; the angular part has been omitted as we constrain ourselves to the PDFs of the e-h pair (exciton). As evident from the Schrodinger equations, we ignored XC (Exchange-Correlations) effects as they hardly contribute to the electronic energies (see SI of methods and calculations).

The confinement potential for the electron between R_1 and R_2 (eq. 5) is calculated by the difference of the electron affinity of the core and the shell. Similarly, for the hole (eq. 6), it is calculated by reducing confinement potential of electrons with the difference of band gap between core and shell[33]. Beyond R_2, the QD is assumed to be in an infinite potential wall, this amounts to a node for the electron and hole wavefunctions at the periphery of the spherical QD. The confinement potential profile of QDs is represented as:

$$V_{e,h} = \begin{cases} 0, & r < R_1 \\ V_{e,h}, & R_2 \le r \le R_1 \\ \infty, & r > R_2 \end{cases} \quad (4)$$

$$V_e = \chi_{CdSe} - \chi_{CdS} \quad (5)$$

$$V_h = E_{bg(CdS)} - E_{bg(CdSe)} - V_e \quad (6)$$

Here, we have assumed the interface between core and shell to be step-like or non-alloyed. Therefore, the confinement potential is assumed constant for the core and shell domains. The surface edge effects and mismatch effects were ignored due to small lattice mismatch (See SI of methods and calculations).

For the interface between the domains, the BenDaniel–Duke boundary conditions[34,35] is applied as:

$$|R_{e,h}(r),\ |_{r=R_1} = |R_{e,h}(r)|_{r=R_2} \quad (7)$$

$$\left|\frac{1}{m_{e,h}^{*(CdSe)}}\frac{dR_{e,h}(r)}{dr}\right|_{r=R_1} = \left|\frac{1}{m_{e,h}^{*(CdS)}}\frac{dR_{e,h}(r)}{dr}\right|_{r=R_2} \quad (8)$$

Equation (7) represents the continuity of the RDF across the interface in the zeroth order differential. The BenDaniel-Duke boundary conditions cause the continuity of the probability current densities at the interface (eq. 8). Here, the effective mass of the charge carriers at core and shell (See SI of Table S1) are different across the interface[36], therefore, the boundary conditions compensates the effective mass mismatch at the interface unlike in the case of classical boundary equations which is introduced for the continuity of the radial wavefunctions of the first order differential.

The electrostatic potentials induced by the electron and hole can be solved by the Poisson equations, as given below[30]:

$$\vec{\nabla}\kappa(r)\,\vec{\nabla}\phi_e = q_e\rho_e \quad (9)$$

$$\vec{\nabla}\kappa(r)\,\vec{\nabla}\phi_h = -q_h\rho_h \quad (10)$$

here, $\boldsymbol{\rho_e(r)}$ and $\boldsymbol{\rho_h(r)}$ are the linear carrier charge densities, and $\boldsymbol{\kappa(r)}$ is the position dependent bulk dielectric constant of the core and shell of the QD. Equations 9 and 10 models the electronic potential images which arise due to surface polarization at the core-shell boundaries. The charge carriers (e–h) linear densities are as follows[33]:

$$\rho_e = \frac{1}{4\pi}q|R_e^{n,l}(r)|^2 \quad (11)$$

$$\rho_h = \frac{1}{4\pi}q|R_h^{n,l}(r)|^2 \quad (12)$$

Here, n represents the principal quantum number and l is the azimuthal quantum number. Since we are concerned with the modelling of s-s transitions for the excitations, therefore we have assumed $n=1$ and $l=0$ for both the electron as well as a hole.

Results and Discussion

The radial distribution functions (RDF) calculated using the Hartree-Fock mean field approximation method represents the spatial distribution of the charge carriers in QDs in steady-state. Figure 2(a–c) shows the behaviour of the distribution of the electron and the hole RDF of CdSe/CdS QD with the increase of the shell thickness from 0.25–3.25 nm with a fixed core radius of 1.75 nm. The electron's spatial distribution is delocalized across the whole domain due to small effective mass which amounts to high kinetic energy, further facilitated by low band offset of the conduction band ($V_c = 0.32$ eV), making delocalization of the electron into the shell easier. The peak of the curve becomes sharp and increases within the QDs as the shell thickness increases to 1.2 nm (Fig. 2(a)). The sharpness of the electron's RDF is independent of the spatial distribution of the hole since hole and electron are co-localized in the same core region (Type-I exciton) till 1.2 nm. The localization of the electron in QDs is governed by the shell size, which acts as a confinement potential of the electron inhibiting the electron's tunnelling probability from the core towards the shell (Fig. 2a). Beyond 1.2 nm shell thickness, electron's energy decreases, as the full-width half-maximum (FWHM) of the electron RDF is reduced (Fig. 2(d)), demonstrated the hole ($+ve$) and electron ($-ve$) energy curves as a function of QD radius, overlapped into the same region to compare the separation of (e-h) energy. Here, the distribution of hole plays an important role. It is shifting towards the shell from the core; this exerts a Coulumbic force on the electron to attract it towards the shell. Also, there is repulsion between the core localized excitons and the shell localized excitons, as the hole are virtually electronically decoupled from each other and separated by an energy gap below a quasi-continuum of shell's energy levels[5,37] and dynamic coulomb blockade process is responsible for the hindering the relaxation of shell localized holes to the core states for sharp confinement potentials[14]. As a result, the electron's spatial distribution is broadened, which causes a reduction in the sharpness of the peak of the spatial distribution of the electron, and the similar trend is followed till 1.8 nm shell thickness.

Upon further increasing shell thickness (above 1.8 nm), the sharpness of the RDF regains because the hole RDF has completely shifted from the core towards shell. Therefore, the Coulomb interaction between the electron and the hole is static due to confinement of hole in shell completely. Due to this, there is a negligible Coulombic effect on the electron, hence it follows general localization of the electron in the entire domain.

Figure 2(b) shows the hole's probability density function, which is shifting from core to the shell, causing the overlap probability (exciton) to shift towards the shell w.r.t shell coverage. Hole's energy monotonically increases with the increase in shell passivation. Since the hole is confined towards the shell making the radial distribution function more sharper thereby enhance the energy of the hole. In the intermediate state, the hole RDF is localized in the core as well as in shell after the shift due to high effective mass amounting to lower kinetic energy, this intermediate region is known as the pseudo-Type-II exciton. This is facilitated further by low valence band confinement potential ($V_h = 0.43$ eV), which makes the hole easier to penetrate towards shell. This shift is instrumental to the reasons for various effects including the excitation's energy variation and specifically in understanding the transition from Type-I to Type-II excitons.

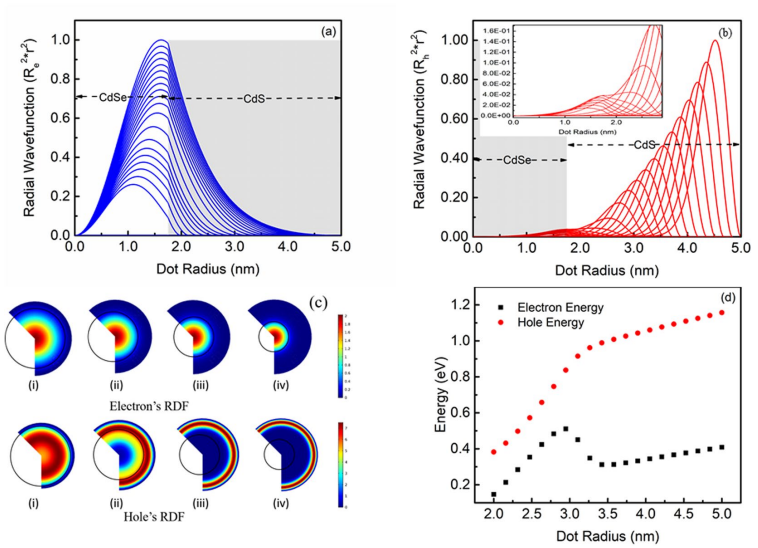


Figure 2. The probability density function of the electron (**a**) and hole (**b**) consists of multiple curves, each RDF curve is plotted as a function of the QDs radius starting from 2 to 5 nm for a single exciton. The inset of (b) shows the magnified curves for the probability density function of the hole in the smaller dot. (**c**) Graphic representation of the electron and hole RDFs of 1.75 nm CdSe core along with a shell thickness of 0.25 (i) 1.25 (ii) 2.25 (iii) and 3.25 nm (iv). The electron's radial wavefunctions are localized over the core domain consistently and the hole's shift towards the shell. (**d**) Energies of the electron and the hole with varying shell thickness.

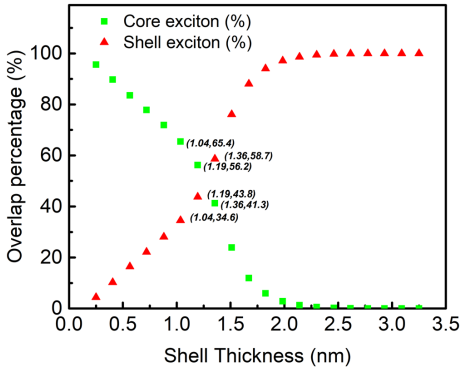


Figure 3. The probability distribution of single exciton (electron-hole) as a function of shell thickness upto 3.25 nm (5.0 nm QD radius) with respect to the percentage area covered by both the core and the shell excitons of the QDs. The number in parenthesis is the shell thickness and overlap percentage.

Relative overlap percentage is a unique way to know the radial density distribution of the exciton over the core-shell to establish the confinement region for excitation energies calculations. The core exciton is decreasing as the thickness of the shell is increased (Fig. 3), since our concern is with the boundary limit 70.7% of core-shell

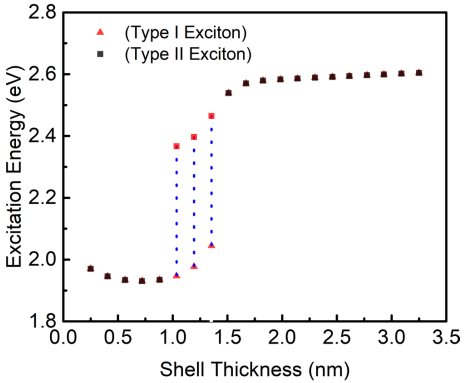


Figure 4. The excitation energies of the excitons with varying shell thickness. The blue vertical dotted lines are representing the dual absorption region.

exciton concentration in core-shell QDs to discriminate the different exciton types. Below 0.88 nm shell thickness, the concentration of the core excitons w.r.t. the total concentration of the excitons is above 70.7% (Type-I exciton). This suggests a shift in the overlap probability of the excitons that can be engineered by applying external electric fields[38].

There are data points between shell thicknesses 1.04 to 1.36 nm (Fig. 3) for which the overlap percentage of the core-shell below 70.7%, which may exhibit dual absorption, two transition energies for each shell thickness suggesting dual absorption from the pseudo-type-II exciton. After this intermediate regime (shell thickness: 1.04–1.36 nm), the confinement region of exciton shifts towards the shell, known as Type-II exciton (Fig. 3). The Type-II exciton produces an absorption which corresponds to the energy difference between core conduction band and the shell valence band (shell to core transitions, Fig. 1).

In the calculation of the excitation energy for dual absorption (See SI of methods and calculations), we took both the bulk bandgaps of the core CdSe and the difference between the conduction band of core and the valence band of the shell (pseudo-Type-II exciton). Similar effects were observed experimentally for the QDs as reported in the literature[13,39,40]. The excitation energy is decreasing with increasing shell width up to 0.72 nm, the decrease is about 3 meV, as shown in Fig. 4. This arises due to the increase in the width of the shell (CdS), also the excitation energy is inversely related to the square of the confinement region's dimension, which agrees to the confinement effect. For 1.04–1.36 nm shell thickness, the curve shows dual excitations with a substantial difference in excitation energy. The overlap percentage of the core or the shell confined exciton w.r.t. each other is below 70.7% in this region. Above 1.67 nm shell thickness, the rate of increase of excitation energy is stagnant after a jump in the excitation energy, due to switching into Type-II exciton having electron delocalized in the core and the hole completely localized in the shell (Fig. 2c). There is no significant increase in the transition energy change w.r.t. shell thickness as the *e-h* pairs are fully confined. The dual emission exhibited by ultra-thick CdSe/CdS core-shell QDs can also be related to this model for further investigations. This effect is primarily due to the thick shell QDs and the nature of the interface between the core and the shell; the step-like confinement potential is more bound to show this effect than the smooth confinement potential[14].

The binding energy of excitons was estimated and are shown in Fig. 5. It has been observed that the values are positive for each shell thicknesses suggesting attractive interactions between the hole and electron. Variation of these interactions are the physical reasons for variation of the binding energy with the shell thickness. The rise in binding energy can be understood predominantly by the increase in the electron and hole probability densities. Therefore, the interactions between them increase. Above 1.36 nm shell thickness, the hole's probability density starts to shift towards the shell from core, since the electron remains confined in the core, therefore, with the increase in probability density of both, the hole-electron separation is accelerated due to the shifting of hole and after the full confinement of the hole, the hole shifts along with the shell thickness increase, which further takes the hole away from the core or here, the core confined electrons. This causes the reduction of the electron and hole Coulombic interaction showing a steep increase in the binding energy[40] (positive slope) till 1.67 nm. After 1.67 nm shell thickness, the separation between them does not increase at a fast rate signifying that the slope of the binding energy becomes less sharp as the shell thickness increases.

The absorption intensity was determined theoretically in terms of the oscillator strength. Figure 6(a,b) gives the normalized oscillator strength of QDs for different shell thickness in the gas phase for both the (a) Type-I (eh|) confinement and (b) Type-II (e|h) confinement. The oscillator strength for second excitation peaks was also estimated for the transitions from 2 S(h) of VB to 1S(e) of CB. The electrons excited from the second sub-band of the valence band (HOMO-1) to the conduction band of the LUMO (Fig. 1). The second excitation state has relatively less oscillator strength due to the negative curve of the hole radial wave function for the second sub-band level of

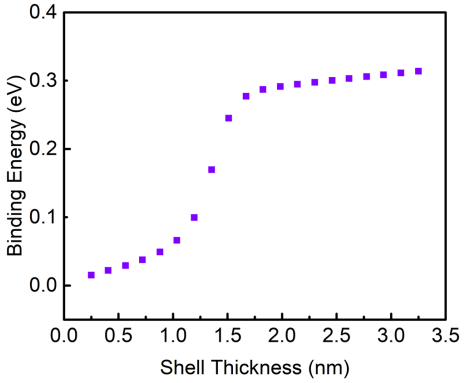


Figure 5. The binding energy of exciton as a function of shell (CdS) thickness.

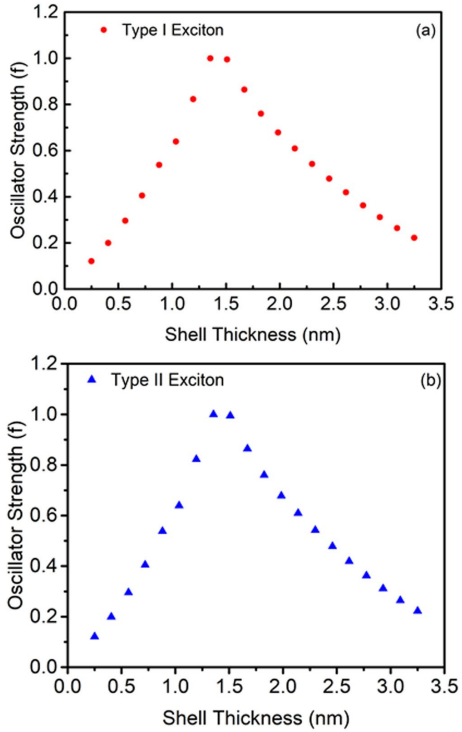


Figure 6. The oscillator strength of Type I (**a**) and Type II (**b**) excitons as a function of the shell thickness in the gas phase.

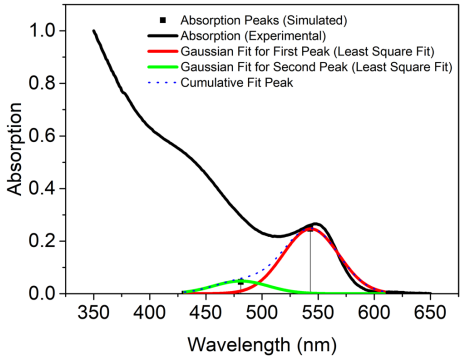


Figure 7. The fitted Gaussian curves on simulated peaks overlapped with the experimental absorption spectrum of CdSe/CdS QDs.

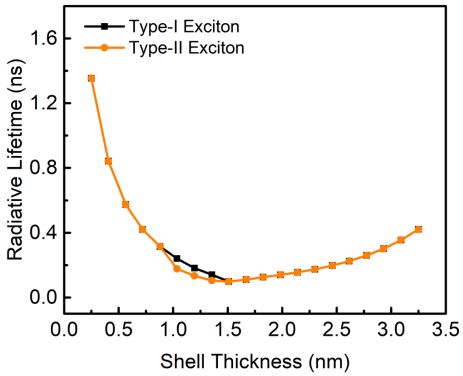


Figure 8. An estimated natural lifetime of excitons (Type I and II) as a function of shell thickness.

the valence band having a single node. The overlap integral of the second excitation is lower than the first excitation peak due to destructive interference (SI of Fig. S1). Thus, the oscillator strength follows the same trend[11]. Further, the simulated relative oscillator strength was compared[12,31] with the normalized experimental absorption spectrum obtained for CdSe/CdS QDs prepared using chemical route at high temperature. The core size of synthesized QDs was typically 1.75 ± 0.25 nm and the shell thickness was 0.25 ± 0.25 nm, which produces the first absorption (excitation) peak at 547 nm[12] and 590 nm[31] and the corresponding simulated value was obtained at 542 nm. The second excitation peak was observed experimentally at 438 nm and theoretically at 481 nm. The difference in the excitation energy might be due to the non-uniformity (shape) of the core-shell structure prepared experimentally, 1-D modelling and also due to the different environments, i.e., the liquid phase for the experimental and gas phase for theoretical.

According to our model, the shell thickness for the particular excitation energy matches with 0.25 nm, which is well within the allowed error estimation of about 0.2/0.15 eV[12,31]. The multiple peak fit is employed to simulate Gaussian curves to represent the calculated excitation peaks and superposed on the actual absorption curve along with the summated Gaussian curve for the first and second excitation bands (Fig. 7). It is important to mention that the modelled QDs follow the similar trend, i.e., the excitation energies decrease with increasing the shell thickness up to 3–4 monolayers[37] (e.g., 1 ML = 0.25 nm).

Figure 8 shows the variation of radiative (natural) lifetime time as a function of shell thickness estimated using the equation provided in supplemental information (Methods and Calculations). As the shell thickness increases from 0.2–1.4 nm, the lifetime decreases and become nearly constant afterwards with slight increase. This trend is quite convincing as per the reported[12,41] experimental values. There are evidences[42,43] for this trend at very low

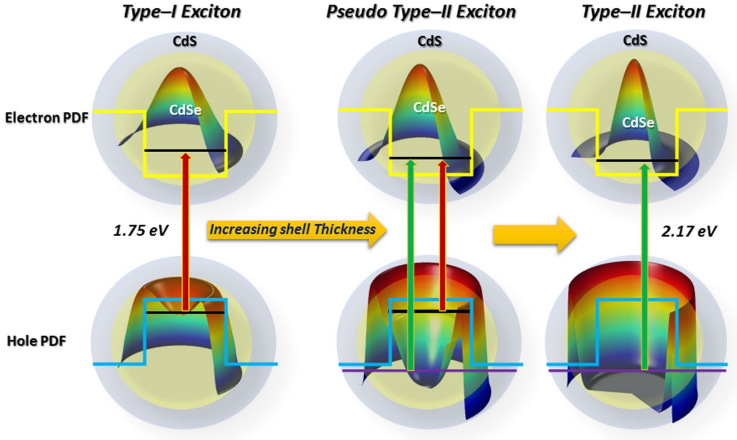


Figure 9. Graphical representation of type I, pseudo type-II and type-II excitons

temperatures, e.g., at about 1.5 K which arises due to negligible non-radiative rates. The inorganic shell passivation leads to decrease in lifetime[44], as the overlap integral between electron and hole increases overall[23], since there is increase in electron's RDF in the core region and the increase in the spread of the hole's RDF in the region of the Type-I QDs. Beyond Type-I QD structure, the increase in shell thickness, there is a decline in the overlap integral of the e-h pair which constitutes for the increase in the radiative lifetime[45]. Many reports[41,42,45] have confirmed that the shell passivation increases the quantum yield of the QDs, which is related to the decrease of the radiative lifetime, as the non-radiative rates becomes less[45]. This is due to, the Auger decay mitigation, as the shell blocks electrons transfer into the dot[46]. Typically, for 0.25 nm shell thickness, the calculated lifetime is 1.35 ns, whereas the radiative lifetime of the experimentally prepared QDs is 0.39 ns[41]/2.18 ns[12]. The difference of the lifetimes is compensated primarily due to the 1-D description used in our modelling, ignoring of non-radiative recombinations, the synthesized QDs may not be perfectly spherical, may contain defect states and the environment (dielectric) effects. Thus, the radiative and non-radiative transitions are quite relevant parameters for the device applications, e.g., solar cells, LEDs, photo-detectors, etc.

Conclusions

Theoretical modelling was employed to obtain the spatial distribution of charge carriers in the core-shell QDs. A charge carries separation between the core confined, and shell confined excitons were obtained above the four monolayers (with a fixed core size QDs). However, beyond six monolayers, the binding and excitation energies are not size-sensitive. In addition to the type-I and type-II excitons, pseudo-type-II excitons (as illustrated in Fig. 9) were obtained within specific shell thickness. The pseudo-type-II excitons lead to the dual absorption bands, which may also give dual emission in the step-like confinement potential[13]. Though the mechanism suggested is debatable and requires much attention on this front for further explanation. We would like that this study will contribute a lot to understanding the basis of electronic and optical properties of the excitons in multi-layered heterostructure QDs and will provide information for material fabrication.

References

1. Rainò, G. *et al.* Probing the wave function delocalization in CdSe/CdS dot-in-rod nanocrystals by time-and temperature-resolved spectroscopy. *ACS Nano* **5**, 4031–4036 (2011).
2. Laheld, U. E. H. & Einevoll, G. T. Excitons in CdSe quantum dots. *Physical Review B* **55**(8), 5184–5204 (1997).
3. Tsuchiya, T. Biexcitons and charged excitons in quantum dots: A quantum Monte Carlo study. *Physica E: Low-Dimensional Systems and Nanostructures* **7**, 470–474 (2000).
4. Pokutnii, S. I. Exciton binding energy in semiconductor quantum dots. *Semiconductors* **44**, 488–493 (2010).
5. Huynh, W. U. *et al.* Charge transport in hybrid nanorod-polymer composite photovoltaic cells. *Physical Review B* **67**, 115326 (2003).
6. Acharya, K. P. *et al.* Linker-free modification of TiO2 Nanorods with PbSe nanocrystals. *The Journal of Physical Chemistry C* **113**, 19531–19535 (2009).
7. Hyun, B. *et al.* Electron injection from colloidal PbS quantum dots into titanium dioxide nanoparticles. *ACS Nano* **2**, 2206–2212 (2011).
8. Acharya, K. P. *et al.* Synthesis of PbS/TiO2 Colloidal heterostructures for photovoltaic applications. *The Journal of Physical Chemistry C* **114**, 12496–12504 (2011).
9. Kumar, S., Jones, M., Lo, S. & Scholes, G. Nanorod heterostructures showing photoinduced charge separation. *Small* **3**, 1633–1639 (2007).
10. Aktürk, A., Sahin, M., Koc, F. & Erdinc, A. A detailed investigation of electronic and optical properties of the exciton, the biexciton and charged excitons in a multi-shell quantum dot nanocrystal. *Journal of Physics D: Applied Physics* **47**, 285301 (2014).

11. Smith, A. M., Lane, L. A. & Nie, S. Mapping the spatial distribution of charge carriers in quantum-confined heterostructures. *Nature Communications* **5**, 4506 (2014).
12. Ratnesh, R. K. & Mehata, M. S. Investigation of biocompatible and protein sensitive highly luminescent quantum dots/nanocrystals of CdSe, CdSe/ZnS and CdSe/CdS. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **179**, 201–210 (2017).
13. Chuang, C., Doane, T. L., Lo, S. S., Scholes, G. D. & Burda, C. Measuring electron and hole transfer in Core/Shell nanoheterostructures. *ACS Nano* **5**, 6016–6024 (2011).
14. Pinchetti, V. *et al.* Effect of Core/Shell interface on carrier dynamics and optical gain properties of dual-color emitting CdSe/CdS nanocrystals. *ACS Nano* **10**, 6877–6887 (2016).
15. Kim, S., Fisher, B., Eisler, H. & Bawendi, M. Type-II quantum dots: CdTe/CdSe (Core/Shell) and CdSe/ZnTe (Core/Shell) heterostructures. *Journal of the American Chemical Society* **125**, 11466–11467 (2003).
16. Segarra, C. *et al.* Piezoelectric control of the exciton wave function in colloidal CdSe/CdS nanocrystals. *The Journal of Physical Chemistry Letters* **7**, 2182–2188 (2016).
17. Garcia- Santamaría, F. *et al.* Suppressed auger recombination in "giant" nanocrystals boosts optical gain performance. *Nano Letters* **9**, 3482–3488 (2009).
18. Garcia-Santamaría, F. *et al.* Breakdown of volume scaling in auger recombination in CdSe/CdS heteronanocrystals: The role of the Core-Shell interface. *Nano Letters* **11**, 687–693 (2011).
19. Eshet, H., Grünwald, M. & Rabani, E. The electronic structure of CdSe/CdS Core/Shell seeded nanorods: Type-I or quasi-type-II? *Nano Letters* **13**, 5880–5885 (2013).
20. Shabaev, A., Rodina, A. V. & Efros, A. L. Fine structure of the band-edge excitons and trions in CdSe/CdS core/shell nanocrystals. *Physical Review B* **86**, 205311 (2012).
21. Jaskolski, W. & Bryant, G. W. Multiband theory of quantum-dot quantum wells: Dim excitons, bright excitons, and charge separation in heteronanostructures. *Physical Review B* **57**, R4237–R4240 (1998).
22. Lo, S. S., Mirkovic, T., Chuang, C., Burda, C. & Scholes, G. D. Emergent properties resulting from type-II band alignment in semiconductor nanoheterostructures. *Advanced Materials* **23**, 180–197 (2010).
23. Kocevski, V., Rusz, J., Eriksson, O. & Sarma, D. D. First-principles study of the influence of different interfaces and core types on the properties of CdSe/CdS core-shell nanocrystals. *Scientific Reports* **5**, 10865 (2015).
24. Liu, H. *et al.* A hybrid quantum-classical model of electrostatics in multiply charged quantum dots. *The Journal of Physical Chemistry C* **121**, 26086–26095 (2017).
25. Royo, M., Climente, J. I., Movilla, J. L. & Planelles, J. Dielectric confinement of excitons in type-I and type-II semiconductor nanorods. *Journal of Physics: Condensed Matter* **23**, 015301 (2010).
26. Schimpf, A. M., Gunthardt, C. E., Rinehart, J. D., Mayer, J. M. & Gamelin, D. R. Controlling carrier densities in photochemically reduced colloidal ZnO nanocrystals: size dependence and role of the hole quencher. *Journal of the American Chemical Society* **135**, 16569–16577 (2013).
27. Schimpf, A. M., Knowles, K. E., Carroll, G. M. & Gamelin, D. R. Electronic doping and redox-potential tuning in colloidal semiconductor nanocrystals. *Accounts of Chemical Research* **48**, 1929–1937 (2015).
28. Balet, L. P., Ivanov, S. A., Piryatinski, A., Achermann, M. & Klimov, V. I. Inverted core/shell nanocrystals continuously tunable between type-I and type-II localization regimes. *Nano Letters* **4**, 1485–1488 (2004).
29. Wang, L. W. & Zunger, A. Pseudopotential calculations of nanoscale CdSe quantum dots. *Physical Review B* **53**, 9579 (1996).
30. Âžahin, M., Nizamoglu, S., Yerli, O. & Volkan Demir, H. Reordering orbitals of semiconductor multi-shell quantum dot-quantum well heteronanocrystals. *Journal of Applied Physics* **111**, 023713 (2012).
31. Li, J. J. *et al.* Large-scale synthesis of nearly monodisperse CdSe/CdS Core/Shell nanocrystals using air-stable reagents via successive ion layer adsorption and reaction. *Journal of the American Chemical Society* **125**, 12567–12575 (2003).
32. Christodoulou, S. *et al.* Synthesis of highly luminescent wurtzite CdSe/CdS giant-shell nanocrystals using a fast continuous injection route. *Journal of Materials Chemistry C* **2**, 3439 (2014).
33. Âžahin, M., Nizamoglu, S., Kavruk, A. E. & Demir, H. V. Self-consistent computation of electronic and optical properties of a single exciton in a spherical quantum dot via matrix diagonalization method. *Journal of Applied Physics* **106**, 043704 (2009).
34. Luo, J., Bester, G. & Zunger, A. Long- and short-range electron-hole exchange interaction in different types of quantum dots. *New Journal of Physics* **11**, 123024 (2009).
35. Peleshchak & Bachynsky Electric properties of the interface quantum dot - matrix. *Condensed Matter Physics* **12**, 215–223 (2009).
36. Barsan, V. & Ciornei, M. C. Semiconductor quantum wells with BenDaniel–Duke boundary conditions: approximate analytical results. *European Journal of Physics* **38**, 015407 (2016).
37. Klimov, V. I. *et al.* Single-exciton optical gain in semiconductor nanocrystals. *Nature* **447**, 441–446 (2007).
38. Mehata, M. S. Enhancement of charge transfer and quenching of photoluminescence of capped CdS quantum dots. *Scientific Reports* **5**, 12056 (2015).
39. Mukherjee, A. & Ghosh, S. Optimum excitation photon energy for CdSe-ZnS core-shell quantum dot based luminescence imaging. *Journal of Physics D: Applied Physics* **45**, 195103 (2012).
40. Sarkar, P., Springborg, M. & Seifert, G. A theoretical study of the structural and electronic properties of CdSe/CdS and CdS/CdSe core/shell nanoparticles. *Chemical Physics Letters* **405**, 103–107 (2005).
41. Ratnesh, R. K. & Mehata, M. S. Synthesis and optical properties of core-multi-shell CdSe/CdS/ZnS quantum dots: Surface modifications. *Optical Materials* **64**, 250–256 (2017).
42. Brovelli, S. *et al.* Nano-engineered electron–hole exchange interaction controls exciton dynamics in core–shell semiconductor nanocrystals. *Nature Communications* **2**, 280 (2011).
43. Javaux, C. *et al.* Thermal activation of non-radiative Auger recombination in charged colloidal nanocrystals. *Nature Nanotechnology* **8**, 206 (2013).
44. Hines, M. A. & Guyot-Sionnest, P. Synthesis and characterization of strongly luminescing ZnS-capped CdSe nanocrystals. *The Journal of Physical Chemistry* **100**, 468–471 (1996).
45. Peng, X., Schlamp, M. C., Kadavanich, A. V. & Alivisatos, A. P. Epitaxial growth of highly luminescent CdSe/CdS core/shell nanocrystals with photostability and electronic accessibility. *Journal of the American Chemical Society* **119**, 7019–7029 (1997).
46. Bae, W. K. *et al.* Controlling the influence of Auger recombination on the performance of quantum-dot light-emitting diodes. *Nature Communications* **4**, 2661 (2013).

Acknowledgements

Author thank the Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Govt. of India for financial support (No. EMR/2016/001110).

Author Contributions

Y.N. designed and performed the quantum chemical calculations and analyzed the data and M.S.M. designed and supervised the research. Both authors discussed the results wrote the paper and commented on the manuscript.

www.nature.com/scientificreports/

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-37676-3.

Competing Interests: The authors declare no competing interests.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019

arXiv:1901.06494v1 [cs.CV] 19 Jan 2019

Writer Independent Offline Signature Recognition Using Ensemble Learning

Sourya Dipta Das[1], Himanshu Ladia[2], Vaibhav Kumar[2], and Shivansh Mishra[2]

[1] Jadavpur University, Kolkata, India
[2] Delhi Technological University, Delhi, India

Abstract. The area of Handwritten Signature Verification has been broadly researched in the last decades, but remains an open research problem. In offline (static) signature verification, the dynamic information of the signature writing process is lost, and it is difficult to design good feature extractors that can distinguish genuine signatures and skilled forgeries. This verification task is even harder in writer independent scenarios which is undeniably fiscal for realistic cases. In this paper, we have proposed an Ensemble model for offline writer, independent signature verification task with Deep learning. We have used two CNNs for feature extraction, after that RGBT for classification & Stacking to generate final prediction vector. We have done extensive experiments on various datasets from various sources to maintain a variance in the dataset. We have achieved the state of the art performance on various datasets.

Keywords: Offline Signature Verification, Convolutional Neural Networks, Ensemble Learning, Deep Learning

1 Introduction

Signature verification is a critical task and many efforts have been made to remove the uncertainty involved in the manual authentication procedure, which makes it an important research line in the field of machine learning and pattern recognition. In offline signature verification, the signature is acquired after the writing process is completed, by scanning a document containing the signature. Defining discriminative feature extractors for offline signatures is a difficult task due to different types of signatures from different origin based on language, country or region . Offline signature verification can be addressed with writer dependent and writer independent approaches. Writer independent scenario is preferable over writer dependent approaches, as for a functioning system, a writer dependent system needs to be retrained with every new writer or signer. For a consumer based system, such as bank, where every day new consumers can open their account this incurs huge cost. Whereas, in writer independent case, a generic system is built to model the discrepancy among the genuine and forged signatures. One of the main challenges for the writer independent offline signature verification task is having a high intra-class variability compared to physical

biometric traits, such as fingerprint or iris. This issue is aggravated with the presence of low inter-class variability when we consider skilled forgeries. In a realistic scenario, during training we only have access to genuine signatures for the users enrolled to the system. During operations, however, we want the system not only to be able to accept genuine signatures, but also to reject forgeries. Lastly, the amount of data available for each user is often very limited in real applications. During the enrollment phase, users are often required to supply only a few samples of their signatures. In other words, even if there is a large number of users enrolled to the system, a classifier needs to perform well for a new user, for whom only a small set of samples are available. In this scenario, our model is based on ensemble model of two deep nets which is trained on two different loss function to learn discriminative features for detection of both person & forgery. With our model, we have achieved high accuracy on large scale cross domain datasets.

2 Previous Work

Different hand crafted features have been proposed for offline signature verification tasks. Examples include block codes, wavelet and Fourier series etc [2]. Some other methods consider the geometrical and topological characteristics of local attributes, such as position, tangent direction, blob structure, connected component and curvature [1], grid based methods [7], methods based on geometrical moments [6], and texture based features [8], Projection and contour based methods [3] are also quite used for offline signature verification as a unsupervised method. Apart from that, some structural methods like graph matching [4] and compact correlated features [5] are also become popular in signature recognition. Convolutional Neural Network [10] have been used to transfer learning of macro & micro features from training on a large dataset to using this model as a feature extractor in a another task by using same network. From the perspective of few shot learning, Siamese networks are popular & effective option for offline signature verification. In a recent work, Convolutional siamese network [9] had been used for offline signature recognition.

3 Our Model Description

We have proposed a ensemble model consists of two deepnet model and we have used Regularized Gradient Boosting Tree (RGBT) as a classifier and finally stacking as a ensemble algorithm to get final prediction vector.

3.1 Preprocessing & Feature Extraction

We have used preprocessing techniques from the paper [10] & then we have used two deep net models to extract the feature from each signature. One deepnet (which is called here as "Signet") model is trained on genuine signatures to learn common descriptive features among the writers and another deepnet model

(which is called here as "Signet-f") is trained on genuine & forged signatures both to learn discriminative features between original and fake one. We pass the signature images through both the models and capture the output of the last layer of the convolutional neural network. This gives us features extracted from signet, and signetf models.

3.2 Model Architecture

We have used same Convolutional Neural Network Architecture mentioned in the paper [10] for both CNNs and used their training procedure & loss functions for respective CNNs on the same dataset to get trained CNNs for feature extraction. Description of the CNN is described in [Fig. 1].

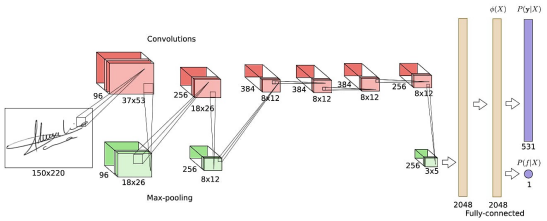


Fig. 1. Description of the CNN architecture used for feature extraction

3.3 Writer Independent Classifier

We have used Regularized Gradient Boosting Tree as our writer independent classifier. It is a sequential technique which works on the principle of an ensemble. It combines a set of weak learners and delivers improved prediction accuracy. At any instant t, the model outcomes are weighed based on the outcomes of previous instant t-1. The outcomes predicted correctly are given a lower weight and the ones miss-classified are weighted higher. Though fitting the training set too closely can lead to degradation of the model's generalization ability. So, to neutralize this effect, we have used regularization methods to reduce this overfitting effect by constraining the fitting procedure. During training, we have set maximum tree depth for base learners as 3, boosting learning rate as 0.1 and number of boosted trees to fit as 100. After getting prediction vector from two classifier, they are fed to stacking module to form a ensemble.

3.4 Stacking as Ensemble Method

Stacking is one of the ensemble model, where a new model is trained from the combined predictions of two or more model. The predictions from the models are used as inputs for each sequential layer, and combined to form a final set of predictions. In the implementation, a logistic regression model is used as the combiner.

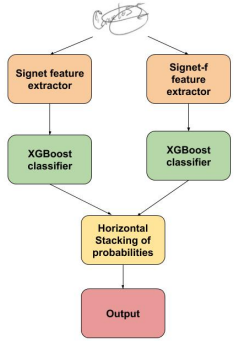


Fig. 2. Signature Verification Model Block Diagram

4 Experimentation

The dataset combining of genuine and forged images is run through the CNN for extracting features. Then, a 33% random test split is used. The 67% data is used for training the RGBT model. Then, from the trained model, the probabilities is calculate for both, the 67% training data and the 33% testing data. The results are evaluated for both, signet and signet-f feature extractors. The corresponding probabilities are also calculated from the features extracted from a signet-f model in the similar fashion. Then, these probabilities are horizontally stacked to create a (n,2) shape feature matrix. This process is done for both, the 67% training data, and 33% test data. A logistic unit is trained on this feature matrix using the 67% training data. And it is tested on the feature matrix of 33% test data. For our implementation & training , we had used a laptop with CPU - Intel(R) Core(TM) i5-6200U CPU @ 2.30GHz and RAM of 8 GB. During the experiments

on various datsets, we observe that the signet model is better at classifying genuine images and the signet-f model is better at classifying forged images. So, it is logical to stack the two classifiers and improve the overall accuracy by a handsome margin.

Table 1. Benchmark Signature Datasets

Dataset Name	Genuine signatures	Fake signatures	Samples in Training Set	Samples in Test Set
CEDAR[11]	1321	1329	1775	875
Hindi-Dataset[8]	3840	4800	5788	2852
Bengali-Dataset[8]	2400	3000	3618	1782

4.1 Results & Discussion

We have evaluated our proposed model with various datasets mentioned in Table 1. We used the performance metric for accuracy of the model described in the paper [9]. Performance with our proposed model and comparison with other state of the art method is described in Table 2. In case of Bengali & Hindi Dataset, our proposed method outperformed other 3 state of the art methods but not be able to outperform Dutta et al., graph matching & SigNet on CEDAR dataset. One possilbe reason for this result may be that for little amount of data, other unsupervised methods and Siamese network (trained on that dataset specifically) are better than ours.

Table 2. Comparison & Performance of the proposed method with other state-of-the-art methods on mentioned signature datasets

Dataset Name	Methods	Accuracy (%)
CEDAR	Word Shape (GSC) (Kalera et al.[2])	78.50
	Zernike moments (Chen and Srihari[13])	83.60
	Graph matching (Chen and Srihari[4])	92.10
	Surroundedness features (Kumar et al.[12])	91.67
	Dutta et al.[5]	**100**
	SigNet[9]	**100**
	Proposed Method	92
Hindi Dataset	Pal et al.[8]	75.53
	Dutta et al.[5]	85.90
	SigNet[9]	84.64
	Proposed Method	**86**
Bengali Dataset	Pal et al.[8]	66.18
	Dutta et al.[5]	84.90
	SigNet[9]	86.11
	Proposed Method	**94**

4.2 Conclusion

In this paper, we have presented a framework based on ensemble learning for offline signature verification. Also, our experiments made on cross domain datasets indicate how well our ensemble model can robustly detect fraudulence of different handwriting style of different signers and forgers with diverse background and scripts. Furthermore, the Ensemble Model designed by us has surpassed the state-of-the-art results on most of the benchmark signature datasets from various scripts & domain. Our future work in this problem will be focus on the multimodal model to incorporate both online & offline signature in one model with different types of data augmentation.

References

1. M. E. Munich, P. Perona, Visual identification by signature tracking, IEEE TPAMI 25 (2) (2003) 200217.
2. M. K. Kalera, S. N. Srihari, A. Xu, Offline signature verification and identification using distance statistics, IJPRAI 18 (7) (2004) 13391360.
3. G. Dimauro, S. Impedovo, G. Pirlo, A. Salzo, A multi-expert signature verification system for bankcheck processing, IJPRAI 11 (05) (1997) 827844.
4. S. Chen, S. Srihari, A new offline signature verification method based on graph, in: ICPR, 2006, pp. 869872.
5. A. Dutta, U. Pal, J. Llados, Compact correlated features for writer independent signature verification, in: ICPR, 2016, pp. 34113416.
6. V. Ramesh, M. N. Murty, offline signature verification using genetically optimized weighted features, PR 32 (2) (1999) 217233.
7. K. Huang, H. Yan, Offline signature verification based on geometric feature extraction and neural network classification, PR 30 (1) (1997) 917.
8. S. Pal, A. Alaei, U. Pal, M. Blumenstein, Performance of an offline signature verification method based on texture features on a large indic-script signature dataset, in: DAS, 2016, pp. 7277.
9. Dey, S., Dutta, A., Toledo, J.I., Ghosh, S.K., Llads, J., & Pal, U. (2017). SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification. CoRR, abs/1707.02131.
10. Luiz G. Hafemann, Robert Sabourin, and Luiz S. Oliveira. 2017. Learning features for offline handwritten signature verification using deep convolutional neural networks. Pattern Recogn. 70, C (October 2017), 163-176. https://doi.org/10.1016/j.patcog.2017.05.012.
11. M. K. Kalera, S. Srihari, A. Xu, Offline signature verification and identification using distance statistics,International Journal of Pattern Recognition and Artificial Intelligence 18 (07) (2004) 13391360. https://doi.org/10.1142/S0218001404003630.
12. R. Kumar, J. Sharma, B. Chanda, Writer-independent offline signature verification using surroundedness feature, PRL 33 (3) (2012) 301308.
13. S. Chen, S. Srihari, Use of exterior contours and shape features in offline signature verification, in: ICDAR, 2005, pp. 12801284.