A MAJOR PROJECT-II REPORT

ON

Detection of Cyberbullying Text Using Hybrid Neural Network Architecture

Submitted in Partial Fulfillment of the Requirement for the Degree of

MASTER OF TECHNOLOGY

in

Computer Science & Engineering

by

Mr. Rishabh Chakraborty
(Roll No-2K22/CSE/19)

Under the Supervision of

Dr. Minni Jain
(Assistant Professor)
Department of Computer Science and Engineering
Delhi Technological University

DEPARTMENT OF COMPUTER SCIENCE AND

ENGINEERINGDELHI TECHNOLOGICAL UNIVERSITY,

(Formerly Delhi College of Engineering)

BAWANA ROAD, DELHI-110042

MAY 2024

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE DECLARATION

I **RISHABH CHAKRABORTY** hereby certify that the Major Project II work which is being presented in the report entitled **Detection of Cyberbullying Text Using Hybrid Neural Network Architecture** in partial fulfillment of the requirements for the award of the Degree of Master of Technology submitted in the Department of Computer Science and Engineering, Delhi Technological University, Delhi. The project work carried out under Dr. Minni Jain's supervision is authentic and hasn't been copied from any source without proper citation.

The matter presented in the report has not been submitted by me for the award of any other degree of this or any other Institute.

Place: Delhi **Rishabh Chakraborty**

Date:

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE

Certified that Rishabh Chakraborty (Roll no 2K22/CSE/19) has carried out his Major Project II research work presented in this report entitled "**Detection of Cyberbullying Text Using Hybrid Neural Network Architecture"** for the award of **Master of Technology** from the Department of Computer Science and Engineering, Delhi Technological University, Delhi under my supervision. The report embodies the results of original work, and studies are carried out by the student himself and the contents do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Place: Delhi
Date:

Dr. Minni Jain
(Supervisor)
Assistant Professor,
Department of CSE,
DTU-Delhi, India

ACKNOWLEDGEMENT

I would like to express my deep appreciation to **Dr. Minni Jain**, Assistant Professor at the Department of Computer Science and Engineering, Delhi Technological University, for his invaluable guidance and unwavering encouragement throughout the course of this research. Her vast knowledge, motivation, expertise, and insightful feedback have been instrumental in every aspect of preparing this research plan.

I am also grateful to **Prof. Vinod Kumar**, Head of the Department, for his valuable insights, suggestions, and meticulous evaluation of my research work. His expertise and scholarly guidance have significantly enhanced the quality of this report.

My heartfelt thanks go out to the esteemed faculty members of the Department of Computer Science & Engineering at Delhi Technological University. I extend my gratitude to my colleagues and friends for their unwavering support and encouragement during this challenging journey. Their intellectual exchanges, constructive critiques, and camaraderie have enriched my research experience and made it truly fulfilling.

While it is impossible to name everyone individually, I want to acknowledge the collective efforts and contributions of all those who have been part of this journey. Their constant love, encouragement, and support have been indispensable in completing this MTech Project II.

Rishabh Chakraborty
(2K22/CSE/19)

ABSTRACT

The rise of digital technology in the modern era and the proliferation of online social media platforms and different online forums have led to unparalleled degrees of communication and sharing of information. Amidst the benefits, it has also led to a pervasive issue in the digital world known as cyberbullying, leading to significant challenges to the well-being of individuals and a threat to societal harmony. This project research work involves a thorough review of the recent advancements in deep learning techniques by various researchers to automate the process of cyberbullying detection. This project work investigates various deep learning techniques like Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), transformers, Graph Convolution Networks (GCN), "Hybrid" based deep models and also a few Machine Learning and AI-based techniques to encounter cyber hate text. The survey is conducted on various private and publicly available datasets to gain insights into these diverse Deep Learning techniques, highlighting their performance, strengths, and limitations. My research experiment reveals that LSTM and Bi-LSTM deep models achieved exceptional performance and BERT, m-BERT and modified BERT models achieved good F-1 scores in detecting toxic content across multiple languages. The hybrid-based models and the introduction of the GCN model are also showing promising results in this domain.

Based on a SWOT analysis approach, this study looks at the phenomenon of cyber hate texts spreading on social sites in depth which will provide valuable insights for researchers, practitioners, and policymakers that will guide in combating cyberbullying detection and the selection and choice of appropriate models**.**

This comprehensive in-depth analysis will provide valuable insights for researchers, practitioners, and policymakers that will guide in combating cyberbullying detection and the selection and choice of appropriate models.

Keywords: Slang words, Multilingual, SWOT Analysis, Transformers, Deep Learning.

LIST OF PUBLICATIONS

1. My paper titled **“Cyber Bullying on Social Media: Comprehensive Review and a SWOT Analysis Approach”** was accepted successfully on 4th May 2024 for publication in the **Springer’s 5th International Conference on Data, Engineering, and Applications (IDEA 2024),** Bhopal, India.

Author Name: Rishabh Chakraborty

Conference Date with Venue: 28th June, 2024, Hybrid Mode.

Paper ID: **716**

Indexed in Scopus

2. My paper titled **“A Comprehensive Analysis of Recent Advancements in Deep Learning based Cyberbullying Text Detection”** was accepted successfully on 6th May 2024 for publication in the “**International Conference on Artificial Intelligence, Machine Learning and Big Data Engineering (ICAIMLBDE)” organized by ISETE, India.**

Author Name: Rishabh Chakraborty

Conference Date with Venue: Presented on 14th May, 2024, Hybrid Mode.

Paper ID: **IST-BDE-PUNE-200524-5758**

Indexed in Scopus

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND NOMENCLATURES

S. No.	Abbreviations	Full Form
1	OSN	Online Social Network
2	SMN	Social Media Network
3	ML	Machine Learning
4	DL	Deep Learning
5	SWOT	Strength, Weakness, Opportunities and Threats
6	CNN	Convolutional Neural Network
7	RNN	Recurrent Neural Network
8	GNN	Graph Neural Network
9	LSTM	Long Short-Term Memory
10	BERT	Bidirectional Encoder Representations from Transformers
11	GPT	Generative Pretrained Transformers
12	MAU	Monthly Active Users
13	WHO	World Health Organization

14	AI	Artificial Intelligence
15	GCN	Graph Convolutional Neural Network
16	BoW	Bag of Words
17	TF-IDF	Term frequency-inverse document frequency
18	Glove	Global Vectors for Word Representation
19	ULMFiT	Universal Language Model Fine-tuning
20	TP	True Positive
21	TN	True Negative
22	FP	False Positive
23	FN	False Negative
24	N	Total Instances which is the summation of TP, TN, FP, FN
25	XAI	Explainable Artificial Intelligence
26	XLM ROBERTa	Cross-Lingual Language Model Robustly Optimized BERT
27	GRU	Gated Recurrent Unit

CHAPTER 1

INTRODUCTION

1.1 A BRIEF OVERVIEW

In the recent era with the rise in digital technology, the demand for online social networking sites (OSNs), forums, and blogging sites has drastically increased among people from different generations. This led to an unprecedented communication level and information sharing over the internet for various purposes like advertisements, news, marketing, business, etc. The users can share and post anything on OSNs at any point in time that can be beneficial but at the same time can also contribute to sharing offensive content that will be a tremendous danger to the balance maintained in society. The number of active users on different OSN sites has increased over the years ranging from 500 million to 3 billion. Figure 1 demonstrates the distribution of the number of users active every month known as Monthly Active Users, "MAU" is shown.

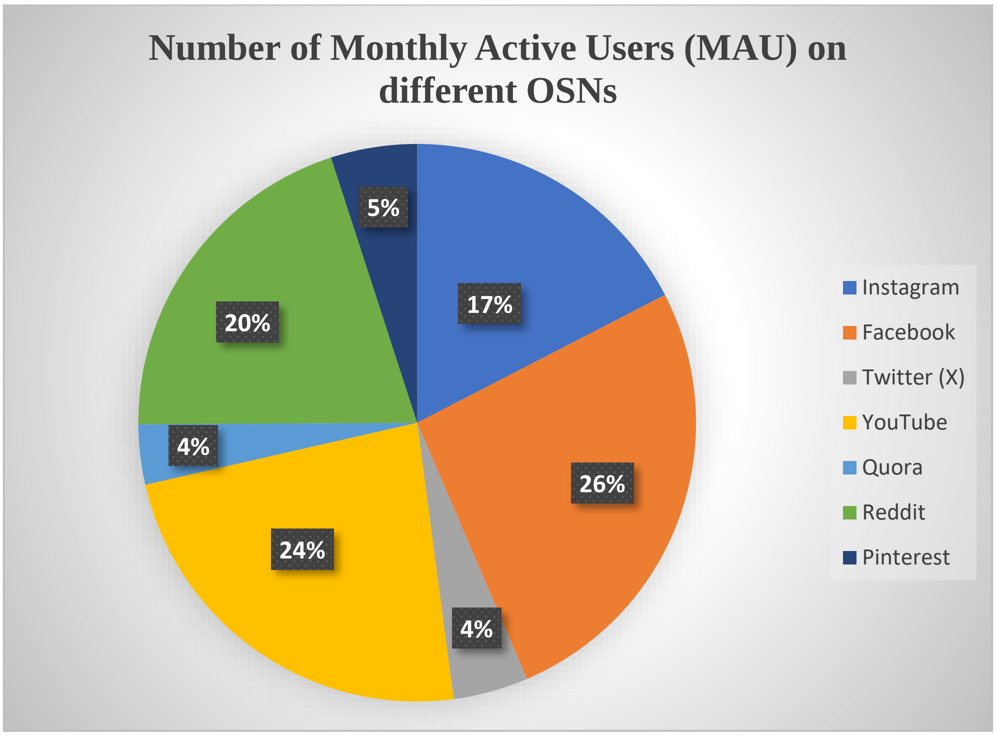


Fig. 1. Monthly Active Users

This rise in the number of users has also given rise to the spread of cyber hate which can be in the form of text, photographs, or videos [3], and among all these, text is the most dominating one in spreading offensive and nasty content. It should be emphasized that there is no formal definition for cyber hate as perspective and context matter [1] which makes it crucial. Bullying that occurs online via text messages, social media, mail, tweets, and other modes of communication via the internet is known as cyberbullying. It is in contrast to traditional bullying and difficult to address and identify as it can occur anytime and anywhere [24, 26] and mostly goes unreported and unseen as the victims can't defend themselves easily. Cyberbullying includes the dissemination of hateful and abusive statements, and making discriminative statements focusing on specific groups or people based on traits like color, religion, sexual orientation, gender, race, and many more giving rise to "cyber hate speech". With the highest number of users on OSNs in India, people are spreading cyberhate in the name of freedom of speech [22], because people are forgetting the thin line between "bully" and "criticize". This is more prominent and present among teenagers and students who frequently enter cyber hate on social media. In India, according to a survey the majority of cyberbullying occurs against famous personalities from diverse industries like Bollywood, Politics, Business, Media, etc. The rise in cyberbullying, fake news, violence, intimidation, sharing propaganda, and nasty content on OSNs has drastically increased in 2024 due to the Indian General Elections, politics being one of the major components of hate speech.

The US Constitution's First Amendment guarantees the right to free expression, which also includes hate speech; however, with some principles, speech that incites violence is deemed to be a serious threat. In India, with the maximum number of users in SMN, people are forgetting the thin line between "criticizing" and "bullying" and hence in the name of "freedom of speech" people are spreading cyber hate. As a result, in SMN many celebrities, politicians, entrepreneurs, teachers, sportspersons, and people from different sectors are often trolled with offensive and foul language. Cyberbullying can lead to a multitude of repercussions that will have a significant negative influence on the physical and mental health of an individual. These include altered sleep patterns, changes in appetite, and physical symptoms like headaches or stomach aches. Moreover, the widespread prevalence of cyberbullying might hamper one's ability to succeed academically or professionally, hence impeding prospects for individual development. According to a report from the World Health Organization (WHO), 4.5% of India's population (~57 million) are depressed and as per the numbers from the World Happiness Report 2024, India ranks 126 out of 146 countries. Cyberbullying is one of the biggest reasons for these numbers as we are living in a virtual world and the opinion of the person in SMN triggers individuals at ease.

1.2 MOTIVATION

The motivation behind the project online cyberbullying text detection on social media using hybrid-neural networks is a response to the rising issue of the proliferation of toxic content on various online platforms. Social media poses serious risks to personal safety and the cohesiveness of society, as it has become a base for cyber hate speech yet provides previously unheard-of levels of connectedness and information exchange. Cyberbullying refers to bullying practices that take place online through text messages, emails, social media, online forums, and other internet-based communication channels. Unlike traditional bullying, cyberbullying is more difficult to recognize and deal with, since it can happen anywhere, at any time, and frequently goes unreported or uses fictitious identities. As per reports from social media, every six out of ten people suffer from cyberbullying once in their life. The main components of cyberbullying are offensive comments, rumors, sexual remarks, trolling, sharing offensive photos/videos without one's consent, objectification, and harassment.

Conventional techniques for identifying cyberbullying frequently fail to capture the nuanced dynamics and complicated dynamics present in the online interactions of various social networking sites. As internet platforms and online forums keep growing and changing with time, so does the potential for cyberbullying, which sometimes goes unreported or receives insufficient attention. This study explores cutting-edge computational methods for cyberbullying detection in light of the pressing need for effective solutions.

The goal of this research study is to understand the importance of the hybrid-based approaches and the effectiveness of cyberbullying detection by utilizing the hybrid neural network design that allows the model to acquire complex language nuances and contextual clues included in cyberbullying literature. By using this strategy, I want to aid in the creation of more resilient and adaptable systems for detecting and stopping cyberbullying activity in online social media platforms.

Furthermore, using the advantages of hybrid neural network techniques like combining traditional deep learning models can effectively encounter cyberbullying speech, ultimately fostering a safer and more inclusive digital environment for all users in OSN sites as it allows the model to acquire complex language nuances and contextual clues included in the cyberbullying domain.

1.3 PROBLEM STATEMENT

With the growing worries about cyberbullying and its negative effects on users' mental health in online social environments, this research work attempts to solve the urgent need for sophisticated computational techniques for the automatic detection of cyberbullying text on various SMNs. The study specifically aims to investigate how well a hybrid neural network architecture can identify texts that constitute cyberbullying. There is a need for more complex strategies since traditional methods frequently fall short of capturing the complex and context-dependent character of cyberbullying behaviors.

Based on the problem statement following questions are identified:

1. What are the various techniques involved in detecting cyberbullying text?
2. What are the recent techniques and advancements?
3. Which technique is more suitable and provides exceptional performance in detecting cyber hate and offensive text?
4. What are the challenges obtained in detecting nasty content?
5. What are the various strengths, weaknesses, and potential threats associated with the detection of cyberbullying text?
6. What are the challenges associated with the dataset in handling the code-mixed data?

With the help of different neural network (NN) architectures, the case study compares the traditional NN approaches and hybrid-based techniques. The work also presents a thorough comprehensive review of previous techniques using various Machine Learning (ML), Deep Learning (DL), and Artificial Intelligence (AI), evaluating their performance based on metrics like precision, recall, accuracy, and F1-score.

This work aims to provide insights into newer ideas and methods used to handle cyber hate. This research work also involves building a SWOT analysis framework using the proposed ideas of various researchers to address the strengths, weaknesses, opportunities, and threats associated with this field.

1.4 WORKING METHOD

This research work method includes access to various databases like “IEEE Explore”, “Scopus”, “ACM”, “Science Direct”, and, “Kaggle” to get the desired articles and the various datasets for carrying forward this research and to understand the different overviews of the datasets used in dealing with cyberbullying speech. The papers selected in my research match the criteria of cyberbullying, cyber hate, and hate or toxic speech in the field of Natural Language Processing. The filtering tool was used to get the most recent and relevant papers in the last seven years in this field; their findings and advancements have been discussed in the former sections.

The subsequent sections of this Major Project II work are divided into several chapters as described below:

Chapter 2- Background

Described about the cyberbullying and cyber hate texts, challenges and techniques to tackle and detect it. Briefly discussed about the preprocessing and datasets used in the later section.

Chapter 3- Literature Survey

Studied the detailed explanation, working of various deep learning methods to tackle these bullying texts. Have done a case study on the different research works in this domain and summarised all the key points of each paper related to the title of the major project.

Chapter 4- Model Description and Comparison

Reviewed the working algorithms and implementation of various hybrid based Deep Learning models used for the detection of cyberbullying texts.

Chapter 5- SWOT Analysis and Result Discussion

A SWOT analysis framework is designed that address the strengths, weaknesses, opportunities, and various threats associated with this field.

Chapter 6- Conclusion

The conclusion of the comprehensive and comparative analysis of numerous models is provided.

CHAPTER 2
BACKGROUND

2.1 Cyber Hate Speech

Spreading of nasty and offensive content, discriminatory remarks, and targeting individuals or a particular community based on various characteristics like ethnicity, race, religion, gender, sexual orientation, disability, physical structure, etc. fall under cyber hate speech. It is to be highlighted that there is no generalized definition of hate speech and hence the context of hate speech on SMN is important.

By generalizing all the definitions for cyber hate speech [1-2], can be defined as "Any form of speech that fosters easily on SMNs that asserts violence, criminal acts and spreading of toxic views towards any individual or groups can be considered as cyber hate speech". The ideology of hate speech is broader due to the constraints in the rapid evaluation of languages, recent trends in languages, and changes in generation. In India, there are 22 official languages as per the 8th schedule of the Indian Constitution and more than 100 unofficial languages. Hence in India, the ideology of hate or non-hate can't be generalized easily and therefore there is a need for proper classifiers to detect the language a proper NLP (Natural Language Processing) based translation scheme is needed.

There are various categories of cyber hate speech and the primary ones are as follows:

1. *Religion:* This category is the most prevalent form and daily, people encounter numerous content on social media spreading hate towards any particular religion. It mainly includes various anti-Hinduism, anti-Islamic, and anti-Christian content, calling for atheism [8] and various propaganda and false information are spread which leads to criminal activities. This type of content is mostly seen on SMN sites before any festival, elections, or protests that create a hostile-like situation.

2. *Gender:* This category includes spreading hate towards any particular gender or sexual orientation of the person. The main victims in this category are people from the LGBTQ+ community. Daily on social media, they receive derogatory comments like g*y, ch**ka, me*t*a, etc that are giving trauma to them. Similarly, females in the name of feminism and freedom of speech are spreading unnecessary hate toward males. After the LGBTQ+ community, the most cyber and offensive hate is spread toward the female gender.

3. *Caste:* This category targets individuals or groups of people based on the caste they belong to. A caste is a form of hereditary, social hierarchical system that is mostly prevalent in South Asian countries. There have always been discriminatory treatments, and

stereotyping towards lower caste people by the members of upper caste. With the influence of social media, hate towards lower caste people has phenomenally increased.

4. Racism: This category encompasses prejudice, discrimination, and structural injustices towards any racial or ethnic group. It includes racial slurs, tribalism, the color of an individual, and prejudice against people from different countries or backgrounds. It can also be categorized under ethnically-based hate speech.

5. Politics: This category involves attacking individuals or groups based on their ideologies, political beliefs, affiliations, religion [16], etc. Hate speech involves the use of aggressive words to disparage a particular person belonging to a political party or political organization. It can take many forms such as taunts, threats, the spread of misinformation, and political riots due to disparaging comments that foster violence or hatred towards different political opponents. It frequently aims to dehumanize people or groups who disagree with the viewpoint of the political person hence leading to dividing the society. In the upcoming Indian General Elections (2024), there is a surge in cyber hate and bullying on social media sites.

6. Physical Structure: This category is more famous in the comment sections of any photos or posts on social media, especially Facebook, Twitter, and Instagram. This includes targeting the characteristics or the physical appearance of human beings. It can be remarking about a person's height, weight, skin tones, facial features or any other attributes related to the physical structure of the body. It can be harmful and foster issues like low self-esteem, emotional distress, depression, and suicidal thoughts in individuals. All these categories are summarized in Table I.

Table I. Various categories

CATEGORIES OF CYBER HATE SPEECH		
Type	***Targeted Community***	***Examples***
Religious	Any religion like Hinduism, Sikhism, Islam, Jainism, Christianity, etc.	Spreading rumors, conversion, terrorism, etc.
Gender	Gay people, straight people, and the other sexually oriented people.	Curse Words
Caste	Mostly lower-caste people.	Curse Words.
Racism	Towards anyone or ethnic groups, mostly tribal people.	Curse Words, Xenophobic.
Politics	Towards any Political Party, person, politician	False propaganda, fake news, curse words.
Physical Appearance	Towards any person especially celebrities and influencers	Fat shaming, Body Shaming, etc.

2.2 Harassment or Bullying

The word bullying is characterized by a pattern of aggressive behaviors that includes repeated acts of harassment, intimidation, or abuse of a person or group to cause them injury, fear, or discomfort. Hate Speech is a component of bullying as hate speech has the potential to create an unfriendly atmosphere conducive to bullying, and those who propagate hate speech may also demonstrate traits associated with bullying. Cyberbullying refers to bullying practices that take place online through text messages, emails, social media, online forums, and other internet-based communication channels. Unlike traditional bullying, cyberbullying is more difficult to recognize and deal with [31], since it can happen anywhere, at any time, and frequently goes unreported or uses fictitious identities. As per reports from social media, every six out of ten people suffer from cyberbullying once in their life. The main components of cyberbullying are offensive comments, rumors, sexual remarks, trolling, sharing offensive photos/videos without one's consent, objectification, and harassment.

2.3 Special Words

The usage of specialized vocabulary, irony, sarcasm, and slang makes it more difficult to understand and identify hate speech in SMNs. These linguistic devices frequently express nuanced ideas that are unclear at first glance. Furthermore, hate speech can pass for irony or sarcasm, making it difficult to recognize without taking context and language indicators into account. Furthermore, hate speech usually uses colloquial or "informal language", necessitating knowledge of a variety of dialects and vernaculars.

Fig. 2. Slang terms used in Modern English

Understanding the context is essential for identifying hate speech because it sheds light on the speaker's motivations, the intended audience, and the underlying message. The context can help to determine if a statement is sarcasm/slang or truly hateful. Furthermore, knowing the audience makes it possible to ascertain whether a speech is offensive to members of a certain social or cultural group. Lastly, examining the larger context of hate speech exposes discriminatory attitudes and unconscious biases, providing insight into the speaker's underlying ideas and objectives.

There are special words in English, shown in Figure 1, that are used daily but the context is important to understand whether it can be categorized as a “hate speech” or a “neutral text”.

According to a report from News18, Lund University in Sweden was heavily trolled, and sexualized comments were made as the first name sounds funny in the Hindi language. There have been many other such instances, where pronunciation or spelling matters a lot. Similarly, the Hindi word “meetha” (sweet in English) and “chakka” (six in English) are Hindi words but they are often used to provide hate towards the LGBTQ+ community and these days most of the people from this community are the victims of these trolls which led to depression and suicide.

The ideology of cyberbullying is it can occur in several forms [23], but it is predominantly in the form of text that involves repeated use of OSN sites and other online platforms to threaten and intimidate individuals. It is different from other forms of bullying as it can be done anonymously and can spread to multiple users over the internet easily.

As per reports from various social networking sites, every seven out of ten people suffer from cyberbullying once in their life and the main target is always the famous personalities and influencers. The primary components of cyberbullying are spreading various rumors, toxic comments, sexual remarks, trolling, sharing offensive photos/videos without one's consent, harassment, and objectification that involves targeted repeated acts.

Cyberbullying is a part of Hate Speech as cyberbullying targets a particular individual repeatedly but cyber hate speech includes individuals or the entire community as a whole based on various characteristics. The basic steps that are involved in dealing with cyber hate detection are shown in Figure 3.

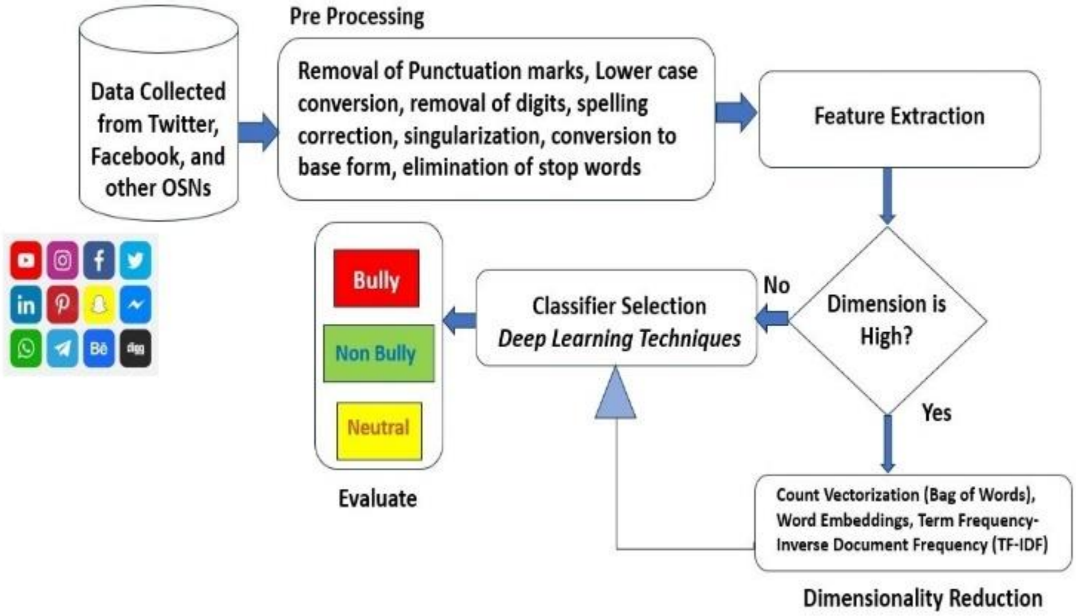

Data Collected from Twitter, Facebook, and other OSNs
Pre Processing
Removal of Punctuation marks, Lower case conversion, removal of digits, spelling correction, singularization, conversion to base form, elimination of stop words
Feature Extraction
Dimension is High?
No
Yes
Classifier Selection
Deep Learning Techniques
Bully
Non Bully
Neutral
Evaluate
Count Vectorization (Bag of Words), Word Embeddings, Term Frequency-Inverse Document Frequency (TF-IDF)
Dimensionality Reduction

Fig.3. Steps in Cyberbullying Detection

2.4 Deep Learning Techniques

Several deep learning-based models have been utilized to deal with cyber hate, that leverages the power of Neural Networks each providing unique advantages and different methodologies to automate the process of identifying and addressing toxic content. Some of the prominent methods used in this field are:

2.4.1 Convolution Neural Networks (CNNs): It was originally developed for image processing tasks, but they have been adapted for text classification tasks [4], which treats text in one-dimensional 1-D space. It can capture the local patterns and hierarchical features in the input text that enable effective identification of cyberbullying text on the OSN sites. An extension to CNN is multi-channel CNN (MC CNN) which includes multiple channels to capture the different aspects of the input text. Every input channel depicts a diverse type of feature representation. The pictorial representation of CNN is shown in Figure 4.

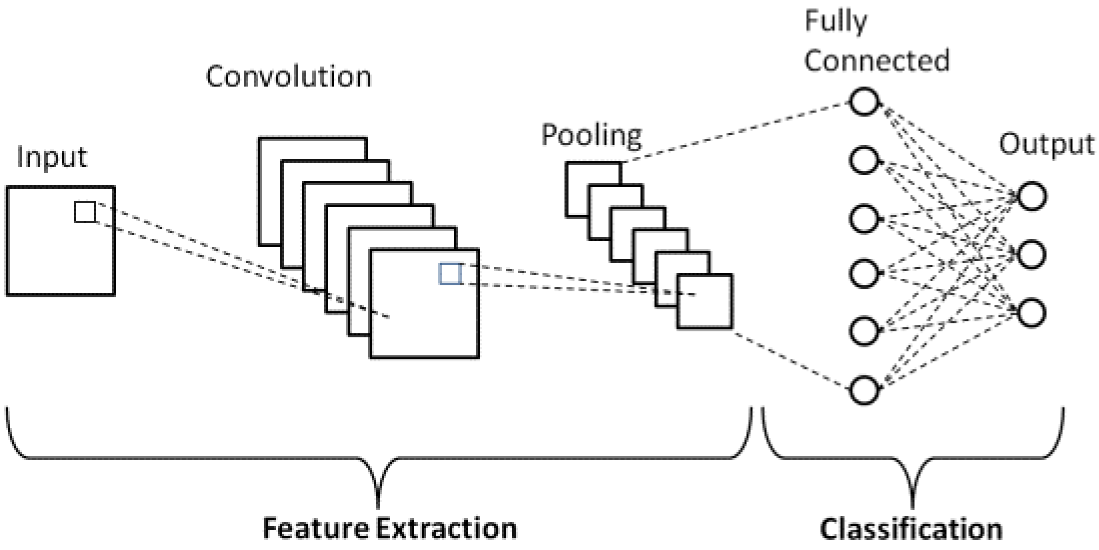


Fig.4. CNN architecture

2.4.2 Recurrent Neural Networks (RNNs): It is designed to handle sequential data like text and hence is the most effective technique in dealing with cyber hate. The main feature of RNN is that it can retain the memory of the previous inputs, making it suitable for long-range interactions in the input text. It can easily capture the nuances of language for gaining insights about the text [12]. It has two special variants known as LSTM and GRU. The architecture diagram of RNN is highlighted in Figure 5.

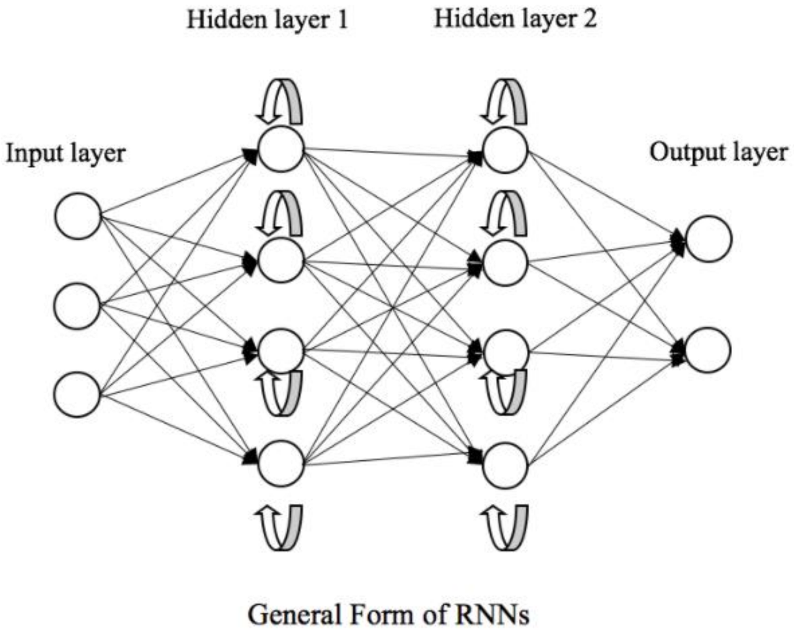


Fig.5. RNN architecture

2.4.3 Long-Short Term Memory Networks (LSTM): It is a subtype of RNN that excels in learning long-range dependencies in textual data. It has memory cells that can store information for a longer period and hence mitigates the problem of “vanishing gradient” faced in the traditional RNN approach. It can learn patterns in abusive textual data and can capture the patterns and context easily in extended conversations [7]. It is useful in sequential tasks in the NLP domain like sentiment analysis, information retrieval, machine translation, etc. There is another variant of LSTM, that is widely proper and is efficient in capturing the context and the semantic meaning of textual content which is BiLSTM.

Bi-LSTM: Bi-directional LSTM stands out as a variant of LSTM architecture extensively applied in NLP, particularly in tasks like Hate Speech Detection, Information Retrieval, Sentiment Analysis, etc. These models process the input sequence in both directions concurrently, in contrast to conventional LSTM models that only process input sequences in one direction (either from past to future or from future to past). Thanks to this bidirectional processing, the model can create a more comprehensive knowledge of the context surrounding each word in the sequence by capturing dependencies from both preceding and subsequent words. By leveraging the information from both preceding and succeeding words in a sentence, Bi-LSTMs excel at capturing the syntactic and semantic structure inherent in language.

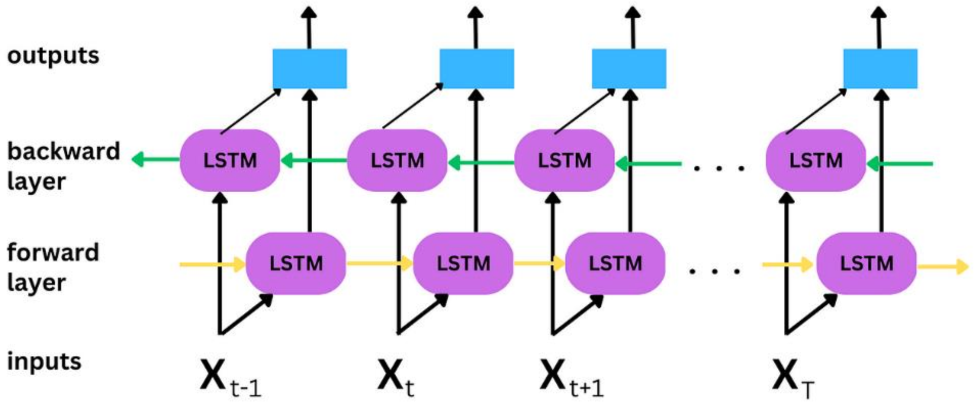


Fig.6. BiLSTM architecture

2.4.4 Gated Recurrent Unit (GRU): It is the simplified version of LSTM with fewer parameters leading to less computation and therefore faster to train and execute. It is more memory-efficient as compared to LSTM. It is helpful when there are limited computational resources on a larger dataset. Similar to BiLSTM, there is **BiGRU** that leverages bidirectional processing that can learn contextual representations easily, incorporating information from both the past and future contexts and helps during complex relationships.

2.4.5 Transformers: It is renowned for its remarkable performance in attention-based mechanisms, capturing global dependencies in the dataset making them widely popular in the domain of NLP tasks. It leverages the ability to capture the context and nuances in the text data [3]. There are numerous pre-trained transformers like Bidirectional Encoder Representations from Transformers abbreviated as BERT, and Generative Pretrained Transformers abbreviated as GPT that are pre-trained on extensive datasets. RoBERTa (Robustly optimized BERT) is a pre-trained transformer developed by a group of researchers in Facebook AI. The use of transformers can help get the semantic and contextual understanding of the text at a deep level as they are pre-trained on a large corpus of datasets.

There are two available variants of BERT:

(1) BERT Base: Supports up to 110 million parameters, and 12 transformer blocks.

(2) BERT Large: Supports up to 340 million parameters and 12 transformer blocks.

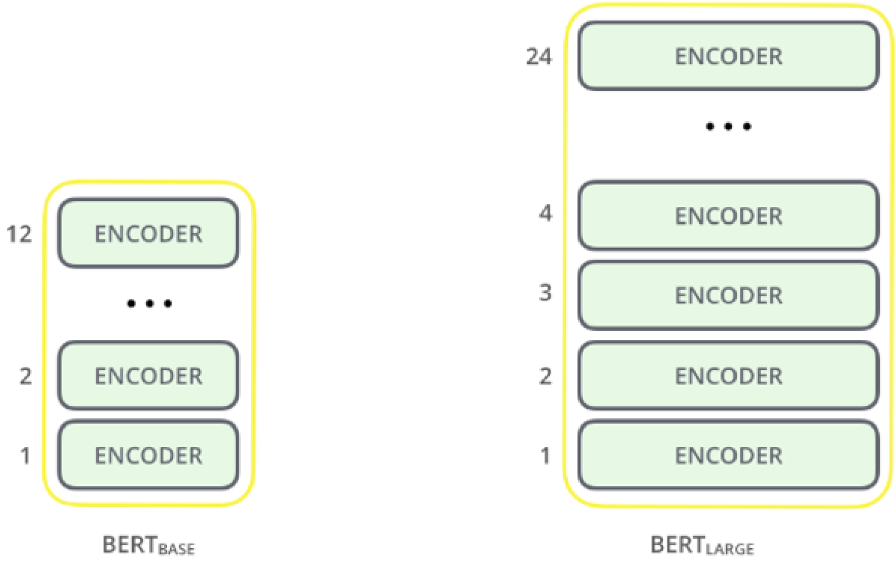


Fig.7. Variants of BERT

Two steps are present in BERT transformer **(a) pre-training and (b) fine-tuning**.

The model is trained on unlabeled data in the pre-training phase on various tasks and all the parameters are fine-tuned via labeled data. The BERT network captures information from both the left and right context of a sentence making it more effective.

2.4.6 Graph Neural Networks (GNN): It is designed to work on graph-structured data (consisting of nodes and edges that represent connections between the nodes). It is helpful to work on complex and relational datasets where every node depicts words in a sentence. It tends to find a pattern in the data that manifests some type of relationship. The majority of GNN networks are Graph Convolutional Networks (GCN) that is similar to CNN and learns features from the neighboring nodes.

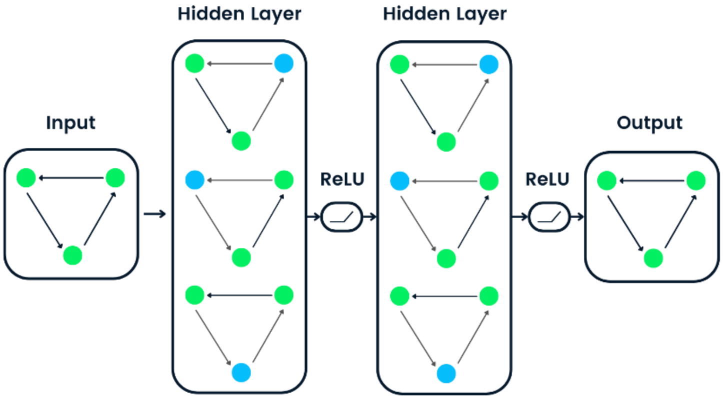


Fig.8. GNN Architecture

2.4.7 Hybrid Model: To enhance the performance of automation of cyberbullying detection, various hybrid models can be built by combining one or more classification techniques. It can include combining diverse DL architectures [5, 8, 15, 19, 20] or integrating various ML techniques with DL or transformers with DL techniques. For example: - Graph Convolutional Networks (GCN) are built by integrating CNN with graph-structured data is a great tool and is widely used by researchers to handle relationship-like structured data [19, 20].

2.5 Dataset

It is the fundamental thing before carrying out any research, to gather the datasets to train a model. There are various types of publicly available datasets on Kaggle and other sites and also private datasets by various authors. Among all these datasets, the majority of the research in the domain of cyber hate speech is carried out in the Davidson-ICSWM [25] dataset which consists of 4163 non-hateful instances and 20620 hateful instances. The dataset has a good sparsity and works pretty well with all DL-based techniques. Similarly, the Waseem-EMNLP [27] dataset consisting of 5850 non-hateful instances and 1059 instances, and the Waseem-NAACL dataset [28] consists of 11501 non-hateful instances and 5406 instances, and VLSP-HSD [29] with 18614 non-hateful and 1731 hateful instances are widely popular in this domain. After going through various research work, it has been observed that most of the datasets are collected crawling from the comment sections of X (Previously Twitter) [5, 7, 9,15] and Facebook.

These datasets are mostly private and not available in the public domain. However, in the future, there is a need for researchers to shift their focus to the comments of other OSN sites as well like Instagram, Quora, Reddit, etc. where online cyberbullying is at its peak. The effectiveness of any model solely depends on the type of language they are trained and also on the training dataset.

The primary issues that might be present in any dataset are as follows:

1. **Imbalance dataset**, which may contain unequal proportions of the vivid classes and categories. For example, if the dataset consists of less cyberbullying speech and more neutral/non-hate text then the result of the model will be biased towards showing neutral results even after having toxic text. To tackle such a problem, there is a need to balance using thorough techniques like sampling, and using of loss functions to ensure accurate and fair prediction by the model.
2. **Sparsity**, means lack of context or categories due to lack of data points in the dataset. The model will fail to predict the rarer and outlier values for sparse data and hence there is a need for additional training and augmentation techniques to handle it.
3. **Cultural and Domain Variation,** the diverse and rich cultures and language variations can create a hindrance to the effective working of the models. For example, the model built with the dataset consisting of English text may not be effective in handling the Arabic text. Hence, adapting the model is important and therefore to different cultural and domain-specific nuances using fine-tuning and multi-tasking approaches.

2.6 Need for Preprocessing

It refers to a set of techniques and vivid operations that are applied to raw data in the dataset. It is done prior to using it for training the model. In general, data preprocessing consists of three basic stages that include ***(a) data cleaning, (b) data integration, and (c) data transformation***. The preliminary process of data to use it for further analysis consists of several steps as discussed below:

1. Removal of noise: It involves the removal of punctuation marks, HTML tags, URLs, hashtag (#), special characters, and non-alphanumeric characters.

2. Normalization: It includes standardization of texts to a consistent format that includes "conversion to lower case", "spelling corrections", "expand contractions (e.g. don't -> do not)", "standard spellings and correction", and "singularization" to mitigate redundancy and variations in the data.

3. Removal of Stop words: The step involves the removal of those words that don't have any semantic meanings (e.g. the, in, is) like articles, conjunctions, and prepositions which helps in dimensionality reduction and provides better analysis.

4. Stemming and Lemmatization: The conversion of words to their root/base form is known as stemming to handle the variations of the same word effectively. It involves removing the suffixes and prefixes and converting them to the root form "stem" (*paying* -> *pay*). On the other hand, lemmatization involves converting a word to the most precise and canonical form based on the closest meaning in the dictionary (*worse* -> *bad*).

Table II shows an example of how the pre-processing stage occurs using an English sentence.

TABLE II. Conversion of data into pre-processed form

Steps	***Original form***	***Pre-processed form***
Removal of punctuation marks	During his 2nd attempting in 2024, the student's hard work was being appreciated @ all his family members.	During his 2nd attempting in 2024 the student's hard work was being appreciated all his family members.
Lowercase Conversion	During his 2nd attempting in 2024, the student's hard work was being appreciated @ all his family members.	during his 2nd attempting in 2024 the student's hard work was being appreciated all his family members.
Deletion of digits	During his 2nd attempting in 2024, the student's hard work was being appreciated @ all his family members.	during his nd attempting in the student's hard work was being appreciated all his family members.
Spelling Correction	During his 2nd attempting in 2024, the student's hard work was being appreciated @ all his family members.	*None in this case* during his nd attempting in the student's hard work was being appreciated all his family members.
Singularization	During his 2nd attempting in 2024, the student's hard work was being appreciated @ all his family members.	during his nd attempting in the student's hard work was being appreciated all his family member.
Convert to Base Form	During his 2nd attempting in 2024, the student's hard work was being appreciated @ all his family members.	during his nd attempt in, the student hard work was be appreciate all his family member.
Stop-Words Elimination	During his 2nd attempting in 2024, the student's hard work was being appreciated @ all his family members.	nd attempt student hard work appreciate family member.

2.7 Feature Extraction and Dimensionality Reduction

It is a crucial step in the process of online cyberbullying text detection using deep learning techniques. It involves transforming text data in raw format into a well-defined structured format that can be utilized by deep learning models more efficiently.

Function: Transforming cleaned text data into numerical features that can be used by Deep learning models.

Correctness: This is a crucial step in the block diagram shown in Figure 3. The methods mentioned (Count Vectorization, Word Embedding, TF-IDF) are appropriate techniques for extracting meaningful features from text. There are various ways to reduce the dimensionality of the text data.

Dimensionality Reduction implies reducing the number of features to the most essential ones.

2.7.1 Bag of Words (BoW)

The model which counts each word's occurrence in a textual document without considering the word's context or sequence, is one of the most straightforward and widely used methods for representing text data. As a result, each page has a "fixed-length vector" with each element representing a "word" in the vocabulary. Although this paradigm is simple to use and effective in capturing word diversity and frequency, it is not without flaws. For data processing, it can produce very huge, sparse vectors that are noisy and inefficient. Additionally, the words' syntactic and semantic meanings are lost, which can be crucial for comprehending the text's structure and meaning.

2.7.2 Term frequency-inverse document frequency (TF-IDF)

It applies a weight to each word depending on the frequency i.e., how often it appears in a text document and how rare it is throughout the entire corpus, which can help get around some of the drawbacks of the BoW paradigm. The concept is that words that appear more often in one document but less frequently in other documents are thought to be more unique and interesting. TF-IDF can increase the relevance of uncommon words that are more precise, like keywords, and lessen the impact of common terms that are not very relevant, such as stop words. For every document, TF-IDF can also produce a vector with a set length, but it can catch more subtleties and changes in the text. It is calculated by the following formula:

$$\mathrm{TF} = \frac{\text{frequency of a term } '\mathrm{t}' \text{ in a document } '\mathrm{d}'}{\text{Total number of terms in document } '\mathrm{d}'}$$

$$\mathrm{IDF} = \log_2(\frac{\text{total documents in corpus}}{\text{number of documents having the term 't'}})$$

$$\mathbf{TF - IDF} = \mathrm{TF\ X\ IDF}$$

2.7.3 Word Embeddings

Using word embeddings which are dense as well as low-dimensional vectors that reflect the syntactic and semantic links among words—is another method to improve the representation of text data and reduce its dimensionality. Neural network models that take advantage of word co-occurrence and context, such word2vec or GloVe, are used to learn word embeddings from vast volumes of text data. Word embeddings preserve the similarity and distance between words in the vector space and can be used to map words with similar meanings or roles to similar vectors. Arithmetic operations on words, such as adding or subtracting vectors to create new words or concepts, can also be made possible via word embeddings.

Word2Vec is a popular word embedding method. It forecasts the context or words that surround a word in a corpus. Two model architectures are provided by Word2Vec: CBOW stands for Continuous Bag of Words and Skip-Gram. Whereas CBOW predicts the target word based on its context, Skip-Gram predicts context words given a target word. Word2Vec generates dense vector representations for words, capturing semantic links and similarities, after training on huge datasets. These embeddings have proven useful in a variety of activities and improve the performance of downstream NLP applications.

Glove embedding is another popular word embedding technique that affects the overall statistical information of a corpus. By examining the worldwide (co-occurrence data) of the words in the corpus, it generates word embeddings. Glove embeddings encode semantic relationship(s) among words by factorizing the word co-occurrence matrix. Due to their aptitude for locating intricate syntactic and semantic traces in the data at hand, they thrive at jobs involving word analogies and semantic similarities. Training classifies general word distributional patterns, producing embeddings in the dataset that accurately capture intricate language relationships.

CHAPTER 3
LITERATURE SURVEY

This section demonstrates a thorough literature review conducted earlier on numerous Deep Learning and hybrid techniques used to encounter cyber bullying texts on social media.
The proliferation of hate speech has been made easier by the rise of social media. Research has looked at how online offensive speech spreads through online channels and the difficulties these platforms have in filtering and controlling it. Scholars also investigate the efficacy of diverse tactics, like community norms and content moderation algorithms, in curbing the dissemination of hate speech on the internet.
However, the community guidelines violations on Facebook, Twitter, and other SMNs are not enough to identify “cyber hate speech” [36] and still daily each of us encounters many bullying and offensive activities taking place due to this there has been a lot of research in this field. As per the data collected by authors in [2], the number of research papers is increasing over the years in this field and the peak value occurred during 2019-20 when COVID-19 started and people were in lockdown. It was the time when everyone was using SMNs and the number of users on these sites drastically increased. Research in this area has a lot of potential because of the surge in both the number of users and data on the internet and the incidence of cyberbullying. The majority of the work in this sector has been done with private data sets obtained via gathering comments from SMNs or with publicly accessible Twitter data sets [37].

1. In the paper by Hind Saleh [3], presents an idea about the usage of pre-trained transformer BERT that leverages effective properties of feature extraction and classification procedures. The author has also done another experiment by performing a comparative study on ML algorithm Logistic Regression (LR) and BiLSTM-based deep model. In this experiment, the author has used Hate Speech Word2Vec [30] as features and a bi-directional LSTM-based deep model as a classifier. The author has concluded that how BERT shows a robust result by combining the benefits of both domain-specific and domain-agnostic word embeddings and saving time and effort to build an embedding model from scratch. However, the proposed second method of domain-specific word embedding has shown better results in detecting hate terms and abbreviations, and intentionally misspelling word meanings over the BERT model.

2. This survey is about two researchers who used the DL technique CNN as a classifier. In this paper by Olumide [4], the authors have 1D-CNN as a classifier and Global Vectors for word representations, GloVe as the word embedding technique for better feature extraction and to capture semantic information from the cyber hate speech. They have compared their model with different ML techniques like Naïve Bayes, Support Vector Machine, Random Forest, and Logistic Regression model along with Bag of Words (BoW) as the word embedding technique. The 1-D CNN technique has improved the overall F-1 score and accuracy as compared to the ML techniques.

In the next paper by Zeleke [6], the authors used an advanced version of CNN i.e., the multi-channel CNN technique abbreviated as MC CNN which is helpful to extract multiple features of input text at the same, since each channel may represent different features or embedding. In this research, the authors have used personalized Amharic 2000 annotated comments from the OSN site and compared their research with the single-channel CNN technique and baseline Support Vector Machine (SVM) technique. The key outcome of the research is that the MC CNN technique outperformed the single-channel CNN technique with an approximate increase of 4% in the F1-score but it under-performs as compared to the baseline SVM model. The primary reason for such a result is that the authors have worked on a smaller dataset due to which ML-based technique has shown better results and this may not be the case with other datasets.

3. In this paper by Dorris [7], the authors have proposed a novel detection model named "HateDefender" using a deep LSTM model and achieved a remarkable average accuracy of 90.8%. The gating signals of LSTM are used to train the input data and capture the nuances of the text. The novel approach uses ranking as the criterion, the gating signals are responsible for computing the word salience rank/score and this score is used to detect which particular word in a text is responsible for causing hate speech.
There are other versions of LSTM models like BiLSTM [14], which is a modified variant of LSTM helpful in producing more accurate output and semantic meaning of the text by combining the layers from both directions of LSTM.

4. In this paper by BR Amrutha [13], the authors have used three different Deep Neural Network (DNN) architectures GRU, CNN, and Universal Language Model Fine-tuning (ULMFiT) to test the detection of cyber hate speech on Twitter-based data. The case study

involves evaluating the performance of the three DNN models. The ULMFit (a type of Neural Network that uses a 3-layer AWD-LSTM), transfer learning-based deep model has shown the highest F1-score of 96.7% in the experiment.

5. In this paper by Faisal [12], the authors worked on the Arabic language dataset collected from comments that are classified into different categories like racism, gender, violence, etc. They proposed deep RNN models DRNN-1 (five hidden layers) used for binary classification and other DRNN-2 (ten hidden layers) for classification. The results are promising that have received an F1-score of 99.73% and 95.38% respectively.

6. These research works use various hybrid models like a combination of various DL based techniques: - CNN+BiGRU [5], GNN+BiRNN & GNN+BERT [8], CNN+LSTM [9], BERT+ANN [11], CNN+GRU [10][15], BERT+GRU [17], BERT+BiLSTM [14] have shown tremendous performance in terms of various performance metrics that can be observed in Table III.

The common understanding from all these experiments is that by combining the features of multiple DL models the overall efficacy increases as it becomes powerful and combines the distinguished features of each DL technique. The hybrid approach produces a robust performance that helps in understanding the semantic nuances of the text more effectively and gives a clear conceptual understanding. Future research can involve the use of a multi-modal hybrid approach that will combine the feature of text and non-text data (images, videos) that will be helpful to prevent all forms of cyber hate prevalent in OSN sites.

7. This survey is about the widely used DL Technique GCN in the span of the last five years by integrating CNN and GNN that is showing promising results to automate the process of detection of cyber hate speech. In the paper by Jason [20], a novel GCN-based architecture is introduced called a Semantic Cosine Similarity Graph Convolutional Network (SOSNet), constructed from thresholded cosine similarities among tweet embeddings.

The effectiveness of any model solely depends on the type of language they are trained on and also on the training dataset. This is because the dataset can consist of imbalance, sparsity, cultural and domain variation, and availability in different languages. So, the authors have made modifications to the dataset using a data mining technique called

Dynamic Query Expansion (DQE) and their study shows the effectiveness of the proposed GCN-based model with the hybrid approach of integrating SOSNET and SBERT, the authors have achieved the maximum accuracy and F1-score.

In the paper by Divyam [18], the authors proposed a novel architecture known as syntax-based LSTM (SyLSTM), which accumulated the syntactic features from the dependency parse tree of a sentence and semantic attributes from word embeddings utilizing GCN. Their approach has significantly outperformed that state-of-the-art BERT model in terms of re-training with few parameters (>110M parameters in BERT and only ~9.5M parameters in SyLSTM). Hence, the computational load has reduced and enhanced the performance of the proposed model. The authors have worked on two instances of the model (i) randomly initialized embedding matrix (SyLSTM) (ii) pre-trained GloVe Twitter embeddings (SyLSTM*). The former one has increased the F1-score by 1.4% which can be observed in Table III.

In this paper by Charles Duong [19], the authors have proposed a novel approach called HateNet consisting of two main components GCN and a weighted drop-edge. To combat the aforementioned challenges in the dataset, the authors have used a similar approach to [18], established a framework for automatic short text data augmentation by forming a semi-supervised hybrid of DQE as explained previously and Substitution based augmentation termed SubDQE. The proposed model has been tested on three different SubDQE augmented datasets and a comparative analysis was done between various tweet embedding methods, baseline data augmentation techniques, and baseline classification models. The SubDQE augmentation helped to improve the classification results and with the HateNet and SBERT configuration outperformed all the other techniques for the three distinct datasets used.

By synthesizing the findings from various research in this field alongside the outcomes collected from my study and understanding, I have compiled them in Table III in the next chapter, providing insights about the comparative advancements of the diverse DL techniques in dealing with cyberbullying texts.

CHAPTER 4

MODEL DESCRIPTION AND COMPARISON

The models and the datasets used along with their performances are summarized in Table III in this chapter. The performance metrics used are as follows:

Metrics	Definition	Formula
Accuracy	It measures the overall correctness of a model's predictions.	$A = \frac{TP + TN}{N}$
Precision	The ability to accurately identify positive instances among all predicted positive instances.	$P = \frac{TP}{TP + FP}$
Sensitivity (Recall)	The ability to correctly identify all actual positive instances among all the positive instances	$R = \frac{TP}{TP + FN}$
F1- score	It is calculated by taking the harmonic mean of precision and recall. It gives a balanced overview of the model's performance.	$F_1 = \frac{2 * P * R}{P + R}$
Macro F1-score	It is calculated by taking the mean of each class F1-score in case binary or multi-label classifications	$mF_1\ score = \sum \frac{F_i}{n}$

The various models and their performances are compared in a tabular format and the models that use hybrid based deep learning techniques are highlighted in bold as shown below:

Table III. Model Analysis

Authors	Dataset	Methods Used	Parameters	Results
Hind Saleh, Areek Alhothali (2023) [3]	Davidson-ICWSM (2017), Waseem-EMNLP (2016), Waseem-NAACL (2017), Balanced Combined (2017)	1. BERT Base 2. BERT Large	F1-Score	1. 0.962 (Davidson), 0.9216 (Waseem_ EMNLP), 0.8472 (Waseem-NAACL), 0.9547 (Balanced Combined) 2. 0.9646 (Davidson), 0.9103 (Waseem_ EMNLP), 0.8521 (Waseem-NAACL), 0.9623 (Balanced Combined)

Olumide Ojo, Thang-Hoang Ta, (2022) [4]	de Gibert et al., ALW 2018	1-D CNN	F1-Score	0.66
Qomarudin Sifak, Erwin (2023) [5]	Twitter Dataset in Indonesian	**A hybrid of CNN and BiGRU and Attention mechanisms** **1. CNN-BiGRU (baseline)** **2. CNN-BiGRU-Attention** **3. CNN-Attention-BiGRU** **4. CNN-Attention-BiGRU-Attention** **5. BiGRU-CNN (baseline)** **6. BiGRU -CNN-Attention** **7. BiGRU - Attention – CNN** **8. BiGRU -Attention-CNN-Attention**	Accuracy	1. 0.8229 2. 0.8341 3. 0.8404 4. 0.8656 5. 0.8781 6. 0.8551 7. 0.8787 8. 0.8812
Zeleke Abebaw, Solomon Atnafu (2022) [6]	Personalized Facebook Comments in Amharic	Multi-Channel CNN (MC CNN)	F1-Score	0.802
Wyatt Dorris, Ruijia (2020) [7]	Twitter Dataset	Deep LSTM	Accuracy	0.9082
Azmine Toushik Wasi (2023) [8]	HateXplain Dataset [10]	**Hybrid of** **1. GNN and BiRNN** **2. GNN and BERT**	Macro F1	1. 0.767 2. 0.797
Pinkesh Badjatiya, Shashank Gupta (2017) [9]	Twitter Dataset	**Hybrid of** **CNN and LSTM**	F-1 Score	0.93
Binny Mathew, Purnyajoy Saha(2020) [10]	HateXplain Dataset	**1. CNN-GRU** **2. BiRNN** **3. Bi-RNN-HateXplain** **4. BERT** 5. **BERT-HateXplain**	Macro F1	1. 0.606 2. 0.575 3. 0.629 4. 0.674 5. 0.687
Harshkumar Mehta, Kalpdrum Passi (2022) [11]	Google Jigsaw, HateXplain	**BERT+ANN**	F-1 Score	0.9414 (HateXplain)
Anezi, Faisal Yousif (2022) [12]	Unique Personalized Dataset of 4203 Arabic comments	Deep RNN (DRNN)	Accuracy	0.9973 (Binary classes), 0.9538 (three classes), 0.8414 (seven classes)
BR Amrutha, KR Bindu (2019) [13]	WikiText103	DNN Architectures 1. GRU 2. CNN 3. ULMFiT	Accuracy	1. 0.9525 2. 0.94 3. 0.967
Saugata Bose, Guoxin Su (2023) [14]	Stormfront (2019)	**BERT+BiLSTM**	F1-score	0.88

Ziqi Zhang, David Robinson (2018) [15]	Twitter Dataset on Refugees and Muslims	**CNN+GRU**	F1-score	0.92
Yingjia Zhao, Xin Tao (2021) [16]	Dravidian Dataset (Code-Mixed comments)	**XLM-RoBERTa + DPCNN**	F1-score	Weighted average F1-score: • Kannada- 0.69 • Malayalam- 0.92 • Tamil- 0.76
Ashfia Jannat, Md. Mohsin (2023) [17]	Bengali Hate Speech Dataset	**GRU+BERT**	Accuracy	0.9556
Divyam Goel, Raksha Sharma (2022) [18]	Davidson-ICWSM (2017) and OLID dataset collected from Tweets [23]	**1. SyLSTM** 2. **SyLSTM***	F1-Score	1. 0.914 2. 0.927
Charles Duong, Lei Zhang (2022) [19]	SubDQE Augmented Dataset 1. HON 2. HANS 3. RSN	**GCN+Weighted Drop-Edge (HateNet)** **Combination of HateNet+SBERT**	Macro F1	1. 0.948 2. 0.973 3. 0.926
Jason Wang, Kaiqun Fu (2020) [20]	DQE Augmented Dataset	**GCN (SOSNet) and SBERT**	F1-score	0.9258
Dipti Mittal, Harmeet Singh [38]	Kaggle-based Hate Speech Dataset	Four models have been used **1. LIME** **2. XGBoost** **3. KTrain** **1. 4. SHAP**	F1-score	1. 0.94 (Class 0) and 0.83 (Class 1) 2. 0.94 (Class 0) and 0.87 (Class 1) 3. 0.98 (Class 0) and 0.79 (Class 1) 4. 0.97 (Class 0) and 0.71 (Class 1)

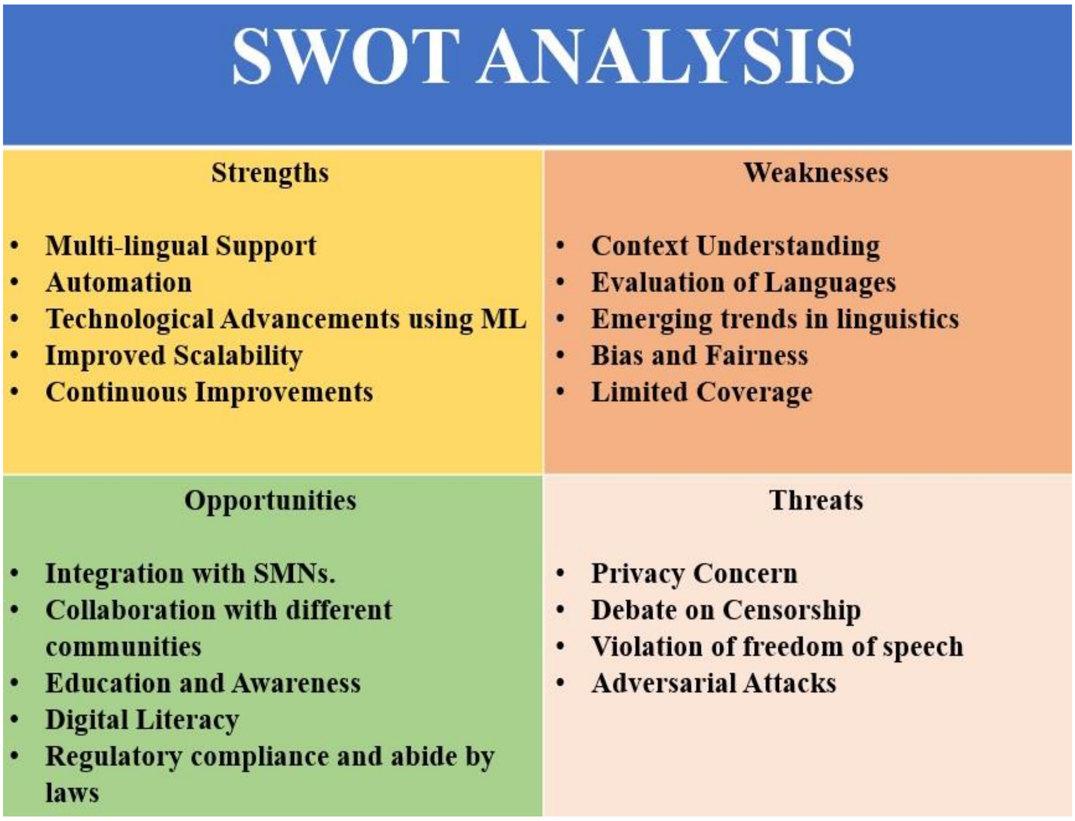
CHAPTER 5
SWOT ANALYSIS AND RESULT DISCUSSION

A strategic planning tool called a SWOT analysis framework is used to determine and evaluate the Strengths, Weaknesses, Opportunities, and Threats associated with a project, company, or any organization.

Making judgments throughout the early phases of the brainstorming process is made easier with the structured framework that SWOT Analysis offers. It offers a methodical framework that facilitates learning at the beginning stages of any decision-making process of a project. Organizations can gain insight into their current position in the market, pinpoint areas for improvement, leverage opportunities, capitalize on strengths, and mitigate weaknesses by conducting a SWOT analysis. They can also prepare contingency plans to address potential threats. Strategic planning, business evaluation, risk management, resource allocation, decision-making, and communication are all aided by the application of SWOT analysis, which aids in the identification of an organization's “internal and external strengths” and “weaknesses” as well as “opportunities and threats”. It also guides decision-making, maximizing strengths, minimizing weaknesses, capitalizing on opportunities, mitigating threats, comprehending the market position, assessing competition, and foreseeing.

SWOT Analysis framework for online cyberbullying detection can be used from various perspectives but in this project work, I am going to shed light on the technical capabilities and what are the internal and external criteria.

A SWOT Analysis framework for the same is shown in the Figure 9, where I brainstormed ideas by understanding the previous methods and the corresponding research gaps in this domain and thereby designed a framework for the associated internal and external criteria that fit within the given context.



SWOT ANALYSIS

Strengths	Weaknesses
• Multi-lingual Support • Automation • Technological Advancements using ML • Improved Scalability • Continuous Improvements	• Context Understanding • Evaluation of Languages • Emerging trends in linguistics • Bias and Fairness • Limited Coverage
Opportunities	**Threats**
• Integration with SMNs. • Collaboration with different communities • Education and Awareness • Digital Literacy • Regulatory compliance and abide by laws	• Privacy Concern • Debate on Censorship • Violation of freedom of speech • Adversarial Attacks

Fig. 9. Cyberbullying Text Detection in SMNs: SWOT Analysis.

5.1 Strengths:

- *Multi-lingual support:* It is one of the major strengths in the automatic detection of cyber hate since the maximum words that are present on the SMNs are in the coded form, a mix of two or more languages. In the paper [32], the authors discussed in India the presence of bilingual and multilingual learning and cross-lingual models [33] can be detected with the help of various Machine Learning Algorithms. Because of its inclusivity, various other detection algorithms like m-BERT, and XLM RoBERTa [35] can be adjusted to fit certain multi-linguistic and cross-linguistic situations, improving their efficacy and accuracy.
- *Automation:* The primary strength of carrying out research in this field is to automate the process because there is humongous data present on different social networking sites and it will be a tedious task to keep humans as moderators and manually remove these offensive texts from these sites. The integration of natural language processing methods with machine learning algorithms to automate the detection procedure. This automation increases productivity by decreasing manual labor and streamlining the identification of cases of hate speech and cyberbullying.

- *Technological Advancements using ML:* Leveraging technological advancements using Artificial Intelligence (AI), XAI and ML algorithms improved the power of decision-making, detection, and classification and overall improved the performance by making the system automated.
- *Improved Scalability:* The data is increasing drastically on SMNs with the increase in the number of users post COVID-19 and these datasets consist of various offensive texts with humongous sizes of tweets and comments collected from different social networking sites as described in [2]. Because of its scalability, detection efforts can effectively handle emerging risks by keeping up with the rapidly increasing volume of online interactions.
- *Continuous Improvements:* The major strength in this field is the scope of continuous improvements. It facilitates a cycle of continual development by illuminating the advantages and disadvantages of current detection techniques. To cope up with the rapidly changing landscape of cyber hate texts and cyberbullying, researchers and developers may update detection models, improve algorithms, and introduce new tactics thanks to this feedback loop.

5.2 Weaknesses:

- *Context Understanding:* As elucidated in the beginning, contextual understanding is the primary concern in the field of NLP. It can be challenging to distinguish between constructive criticism and harmful words without a thorough comprehension of the context, which can result in inaccurate detection as described in [1].
- *Evaluation of Languages:* The majority of challenges occur due to the evaluation of languages with the changing generation and the thought process of individuals in terms of linguistic literacy. A survey found that many words in English that were widely used back in the mid-1900s are not used currently in any conversation. It may be difficult to appropriately assess languages with complicated linguistic systems or those that are less widely spoken. Languages with insufficient training data and lack of formal dictionary meaning or linguistic resources may reduce the effectiveness of various detection algorithms, which would lead to lower detection accuracy and reliability.
- *Emerging Trends in Linguistics*: The current trend of the GenZ generation is to use slang terms, sarcasm, and different ironic words in their words which could be challenging to detect and identify whether it is offensive or neutral. It is again interlinked with the contextual understanding of the text.

- *Bias and Fairness:* It may unintentionally reinforce unfairness and bias. This is especially true if the ML or Deep Learning algorithms are not made to take into consideration a variety of linguistic and cultural settings, and demographic groups or if the training data is biased. This may lead to a disproportionate focus on particular communities or groups, so escalating already-existing disparities. In subtle the algorithm may be more successful at identifying cyber hate directed towards specific racial or religious groups but less successful at identifying hate speech directed towards other groups if the training data is predominately composed of cyber hate aimed towards those groups.
- *Limited Coverage:* The prime weakness is that coverage is limited since these days there are numerous SMNs and blogging sites like Reddit, Quora, and several other apps and websites. It will be a tedious task to collect data from each and everywhere. Access to resources and data may limit detection attempts, resulting in coverage gaps and an incomplete picture of the issue.

5.3 Opportunities:

- *Integration with Social Media Networks (SMNs):* Collaboration with different social media networks presents a chance to directly include hate speech and cyberbullying detection systems into their platforms. By integrating detection algorithms into SMNs' infrastructure, it will become feasible to recognize and remove this hazardous and offensive information instantly, giving users access to a more secure and encouraging online community.
- *Collaboration with different communities:* Social media sites consist of people from diverse backgrounds, cultures, races, and religions. Hence, collaborating with different government and non-government bodies offers the chance to acquire knowledge, viewpoints, and contextual understanding from a range of stakeholders. It is possible to create more complex and culturally sensitive methods for detecting hate speech and cyberbullying as well as to encourage inclusivity and diversity in detection efforts by interacting with other communities.
- *Education and Awareness:* Education and awareness programs can be conducted with the help of various campaigns and the help of podcasts and social media influencers not to promote cyberbullying as it can be a punishable offense. Also, people can be empowered to identify and successfully respond to hate speech and cyberbullying by giving them resources for addressing these issues and increasing knowledge about the impact of harmful internet content.

- *Digital Literacy:* It is highly correlated with the education and awareness portion. Many people are new to SMNs and they do not abide by the laws and different agreements present on these sites. Fostering critical thinking and digital literacy among users—especially youth and teenagers offers a chance to lessen the prevalence and effects of cyberbullying. People can be less vulnerable to online harassment and manipulation and help create a more constructive and upbeat online culture by giving them the information and abilities to use digital spaces securely and responsibly.
- *Regulatory compliance and abiding by laws:* By improvising education, digital literacy, and awareness campaigns governments and internet platforms can work together to develop laws and community guidelines that combat cyberbullying while upholding people's right to privacy and freedom of speech. A more secure and responsible online environment for all users can be achieved by enacting explicit regulations, legal measures, and enforcement strategies to counteract bad information.

5.4 Threats:

- *Privacy Concern:* There can be serious privacy issues while dealing with various data from internet sites. There will be gathering and processing of personal information. These systems run the risk of violating people's right to privacy by tracking their online actions, reading through their correspondence, and maybe creating a profile of them based on private data. Privacy concerns are further exacerbated by the fact that managing and storing vast amounts of user data carries a risk of data breaches, misuse, and unauthorized access.
- *Debate on Censorship:* There is a humongous debate on social media on what exactly hate and nasty content is due to the violation of policy. There is no generalized definition of the term “hate speech” as discussed in the earlier chapter, so it becomes a hot topic for debate due to the different perspectives of different people towards offense and arrogance. Additionally, the tools for automatic detection pose a threat to the users in a way that they might carry away their right to freedom of speech, disagreeing with opinions and neutral or controversial statements will be termed as cyber hate, leading to unfair censorship [34] and it will carry away the basic rights of people to share their points of view online.
- *Violation of Freedom of Speech:* As discussed in the previous threat, these automatic tools to remove hateful and nasty content will carry away the freedom of sharing one’s opinions on SMNs. If the freedom of speech is taken away, it will lead to monopoly giving rise to dictatorship in every sector as the viewpoint of others will be mitigated as described in [34].

It will become easier for the higher authority people to brainwash others leading to a disbalance in peace and harmony in the society.

- *Adversarial Attacks:* These involve creating content that purposely bypasses or causes confusion to the automated system for cyber hate detection. It involves creating fake examples that may look normal to humans but can trick the machine learning models by taking advantage of flaws in the system. These attacks will reduce the efficiency and overall accuracy of the model and can take many different forms like data poisoning, corrupting the data, and evasion.

CHAPTER 6
CONCLUSION

This comprehensive analysis highlights the significant advancements leveraging diverse Deep Learning techniques to encounter the rise in the usage of nasty and offensive language over social networking sites. There has been extensive and similar research in this field, but still, it will remain a new research topic in the computing domain due to the evaluation of languages and linguistic trends with the changing generations. Hence, there is a need for constant updates because the model used for one dataset may not effectively work with another dataset with the change in patterns of linguistics. So, researchers need to be constantly updated with the recent advancements and progress made in the domain of automation of cyberbullying speech. This research also leverages the strengths and weaknesses of the diverse deep models in the detection of toxic content Amidst these obstacles, there are promising avenues and in recent times, deep learning and XAI-based models have emerged as promising tools to automate the detection of cyber hate. Among all these models, the survey enlightens that the different variants of LSTM, BiLSTM, and deep LSTM models have shown significantly good performance in terms of accuracy and F-1 score, and leveraging hybrid and modified BERT models has shown tremendous results with multilingual and cross-lingual datasets. The advancements using the GCN classifiers have a tremendous scope for future research.

The objective of my research is to provide valuable insights into the recent advances in the domain of cyberbullying using various DL and hybrid techniques. The research work of multiple authors has been appraised. This analysis will serve as a valuable resource for multiple researchers, policymakers, stakeholders, and practitioners for the model selection and make improvements in their algorithms to enhance the efficiency of automation fostering a safer environment on social networking sites.

REFERENCES

[1] S. MacAvaney, H.-R. Yao, E. Yang, K. Russell, N. Goharian, and O. Frieder, "Hate speech detection: Challenges and solutions," PLoS ONE, vol. 14, no. 8, p. e0221152, 2019. DOI: 10.1371/journal.pone.0221152.

[2] F. Alkomah and X. Ma, "A Literature Review of Textual Hate Speech Detection Methods and Datasets," Information, vol. 13, p. 273, 2022. DOI: 10.3390/info13060273.

[3] S. Saleh, H. Alhothali, A. Moria, and K. Moria, "Detection of Hate Speech using BERT and Hate Speech Word Embedding with Deep Model."

[4] Ojo, Olumide & Ta, Thang & Gelbukh, Alexander & Calvo, Hiram & Sidorov, Grigori & Adebanji, Olaronke. (2022). "Automatic Hate Speech Detection Using CNN Model and Word Embedding." Computación y Sistemas. 26. 10.13053/cys-26-2-4107.

[5] Sifak, Qomarudin & Setiawan, Erwin. (2023). Hate Speech Detection using CNN and BiGRU with Attention Mechanism on Twitter. 170-175. 10.1109/COMNETSAT59769.2023.10420628.

[6] Abebaw, Zeleke & Rauber, Andreas & Atnafu, Solomon. (2022). Multi-channel Convolutional Neural Network for Hate Speech Detection in Social Media. 10.1007/978-3-030-93709-6_41.

[7] W. Dorris, R. R. Hu, N. Vishwamitra, F. Luo, and M. Costello, "Towards automatic detection and explanation of hate speech and offensive language," IWSPA 2020 - Proc. 6th Int. Work. Secur. Priv. Anal., pp. 23–29, 2020.

[8] Wasi, A. T. (2023). Explainable Identification of Hate Speech towards Islam using Graph Neural Networks. *ArXiv*. /abs/2311.04916.

[9] P. Badjatiya, S. Gupta, M. Gupta, and V. Varma, "Deep learning for hate speech detection in tweets," *26th Int. World Wide Web Conf. 2017, WWW 2017 Companion*, no. August 2020, pp. 759–760, 2019.

[10] Binny Mathew, Punyajoy Saha, Seid Muhie Yimam, Chris Biemann, Pawan Goyal, and Animesh Mukherjee. Hatexplain: A benchmark dataset for explainable hate speech detection. In AAAI Conference on Artificial Intelligence, 2020.

[11] H. Mehta and K. Passi, "Social media hate speech detection using explainable artificial intelligence (XAI)," Algorithms, vol. 15, no. 8, p. 291, Aug. 2022. https://doi.org/10.3390/a15080291.

[12] Anezi, Faisal Yousif Al. 2022. "Arabic Hate Speech Detection Using Deep Recurrent Neural Networks" Applied Sciences 12, no. 12: 6010. https://doi.org/10.3390/app12126010.

[13] Amrutha, B & K R, Bindu. (2019). Detecting Hate Speech in Tweets Using Different Deep Neural Network Architectures. 923-926. 10.1109/ICCS45141.2019.9065763.

[14] S. Bose and G. Su, "Deep one-class hate speech detection model," in Proceedings of the Thirteenth Language Resources and Evaluation Conference, Marseille, France pp. 7040–7048, 2022.

[15] Zhang, Ziqi & Robinson, D. & Tepper, Jonathan. (2018). Detecting hate speech on Twitter using a convolution-GRU based deep neural network.

[16] Y. Zhao and X. Tao, "ZYJ123@DravidianLangTech-EACL2021: Offensive language identification based on XLM-RoBERTa with DPCNN," in Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages, pp 216–221,2021.

[17] Keya, Ashfia & Kabir, Md & Shammey, Nusrat Jahan & Ph. D., M. & Islam, Dr. MD Rashedul & Watanobe, Yutaka. (2023). G-BERT: An Efficient Method for Identifying Hate Speech in Bengali Texts on Social Media. IEEE Access. PP. 1-1. 10.1109/ACCESS.2023.3299021.

[18] Goel, Divyam, and Raksha Sharma. "Leveraging dependency grammar for fine-grained offensive language detection using graph convolutional networks." arXiv preprint arXiv:2205.13164 (2022).

[19] C. Duong, L. Zhang and C. -T. Lu, "HateNet: A Graph Convolutional Network Approach to Hate Speech Detection," 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 2022, pp. 5698-5707, doi: 10.1109/BigData55660.2022.10020510.

[20] Wang, Jason et al. "SOSNet: A Graph Convolutional Network Approach to Fine-Grained Cyberbullying Detection." *2020 IEEE International Conference on Big Data (Big Data)* (2020): 1699-1708.

[21] Marcos Zampieri, Shervin Malmasi, Preslav Nakov, Sara Rosenthal, Noura Farra, and Ritesh Kumar. 2019. Predicting the type and target of offensive posts in social media. arXiv preprint arXiv:1902.09666.

[22] S. Basu and S. Sen, "Silenced voices: unravelling India's dissent crisis through historical and contemporary analysis of free speech and suppression," Information & Communications Technology Law, vol. 33, no. 1, pp. 42–65, 2024. DOI: 10.1080/13600834.2023.2249780.

[23] A. Rajeevan and N. Krishnaraj, "Detection Of Cyberbullying based On Online Social Networks," 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2023, pp. 1-4, doi: 10.1109/ICCCI56745.2023.10128506.

[24] P. Burnap and M. L. Williams, "Cyber hate speech on Twitter: an application of machine classification and statistical modeling for policy and decision making," Policy Internet, vol. 7, no. 2, pp. 223–242, 2015.

[25] T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated hate speech detection and the problem of offensive language," in Proceedings of the 11th Conference on Web and Social Media, AAAI, 2017.

[26] N. Bölücü and P. Canbay, "Hate Speech and Offensive Content Identification with Graph Convolutional Networks," 2021.

[27] Z. Waseem, "Are you a racist or am I seeing things? Annotator influence on hate speech detection on Twitter," in Proceedings of the First Workshop on NLP and Computational Social Science, 2016, pp. 138–142.

[28] Z. Waseem and D. Hovy, "Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter," in Proceedings of the NAACL Student Research Workshop, 2016, pp. 88-93.

[29] Vu, X., Vu, T., Tran, M., & Nguyen, H. T. (2020). HSD Shared Task in VLSP Campaign 2019:Hate Speech Detection for Social Good. ArXiv. /abs/2007.06493.

[30] Gupta, S., and Z. Waseem. 2017. A comparative study of embeddings methods for hate speech detection from tweets. Association for Computing Machinery.

[31] P. Burnap and M. L. Williams, "Cyber hate speech on Twitter: an application of machine classification and statistical modeling for policy and decision making," Policy Internet, vol. 7, no. 2, pp. 223–242, 2015.

[32] K. Sreelakshmi, Premjith B., and Soman Kp, "Detection of Hate Speech Text in Hindi-English Code-mixed Data," Procedia Computer Science, vol. 171, pp. 737-744, 2020. DOI: 10.1016/j.procs.2020.04.080.

[33] S. Gaikwad et al., "Cross-lingual offensive language identification for low resource languages: The case of Marathi," in Proceedings of RANLP, 2021.

[34] T. Gelashvili and K. A. Nowak, "Hate Speech on Social Media," Lund University, 2018.

[35] A. Conneau et al., "Unsupervised Cross-lingual Representation Learning at Scale," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 2020, pp. 8440-8451.

[36] Z. Talat and D. Hovy, "Hateful symbols or hateful people? predictive features for hate speech detection on twitter," in North American Chapter of the Association for Computational Linguistics, 2016.

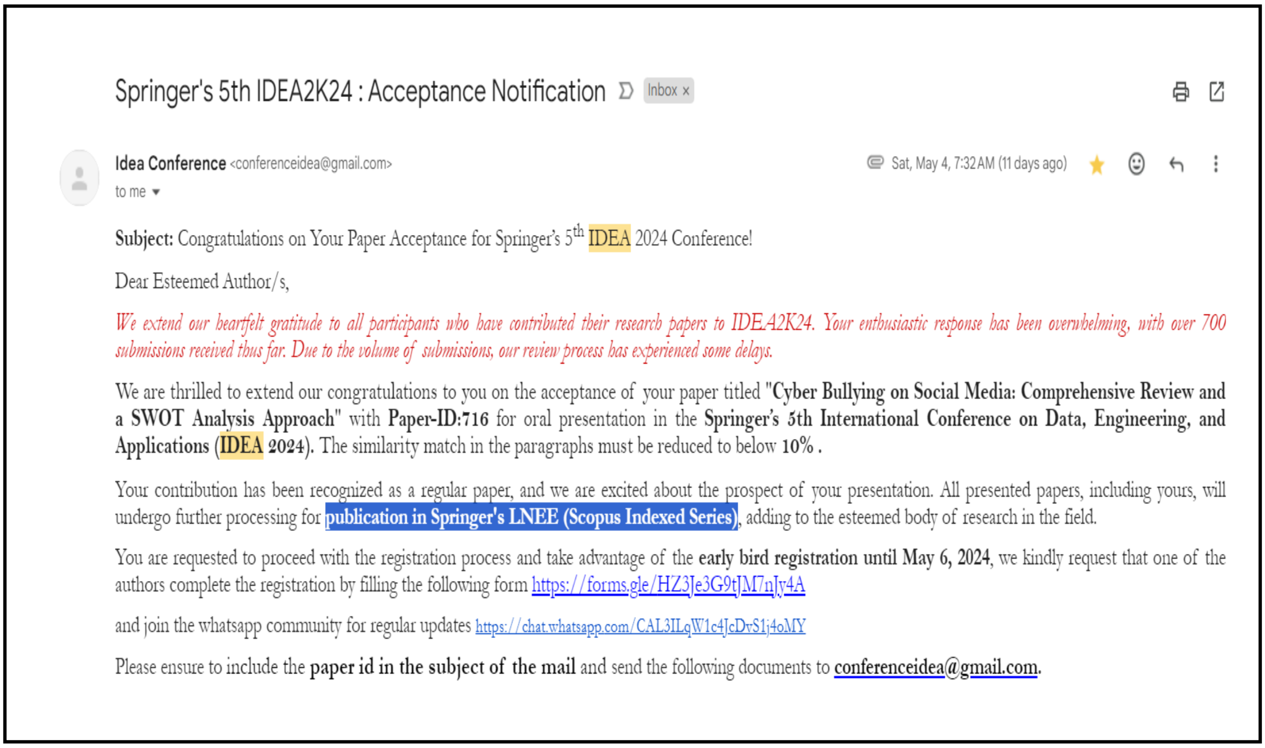
[37] Z. Zhang, D. Robinson, and J. Tepper, "Detecting Hate Speech on Twitter Using a Convolution-GRU Based Deep Neural Network," in Proceedings of the Semantic Web. ESWC 2018, A. Gangemi et al., Eds. Springer, Cham, 2018, vol. 10843. DOI: 10.1007/978-3-319-93417-4_48.

[38] D. Mittal and H. Singh, "Enhancing Hate Speech Detection through Explainable AI," in 2023 3rd International Conference on Smart Data Intelligence (ICSMDI), Trichy, India, 2023, pp. 118–123. DOI: 10.1109/ICSMDI57622.2023.00028.

LIST OF PUBLICATIONS AND THEIR PROOFS

Paper 1

Acceptance Notification



Springer's 5th IDEA2K24 : Acceptance Notification

Idea Conference <conferenceidea@gmail.com>
to me

Sat, May 4, 7:32 AM (11 days ago)

Subject: Congratulations on Your Paper Acceptance for Springer's 5th IDEA 2024 Conference!

Dear Esteemed Author/s,

We extend our heartfelt gratitude to all participants who have contributed their research papers to IDEA2K24. Your enthusiastic response has been overwhelming, with over 700 submissions received thus far. Due to the volume of submissions, our review process has experienced some delays.

We are thrilled to extend our congratulations to you on the acceptance of your paper titled "**Cyber Bullying on Social Media: Comprehensive Review and a SWOT Analysis Approach**" with **Paper-ID:716** for oral presentation in the **Springer's 5th International Conference on Data, Engineering, and Applications (IDEA 2024)**. The similarity match in the paragraphs must be reduced to below **10%** .

Your contribution has been recognized as a regular paper, and we are excited about the prospect of your presentation. All presented papers, including yours, will undergo further processing for **publication in Springer's LNEE (Scopus Indexed Series)**, adding to the esteemed body of research in the field.

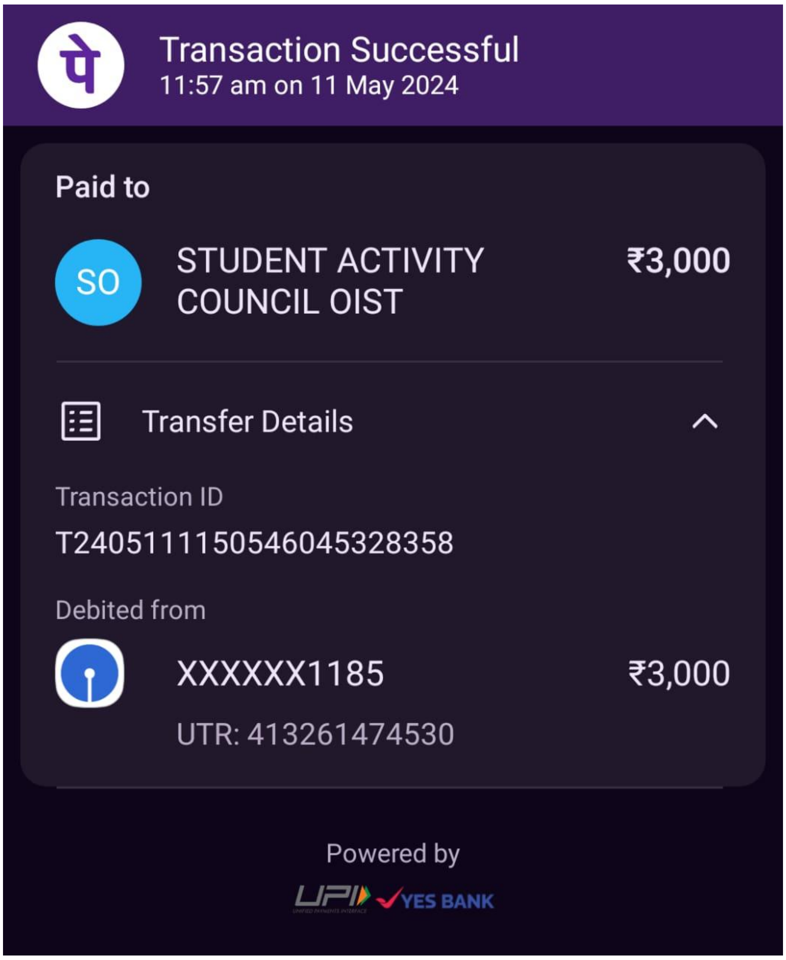
You are requested to proceed with the registration process and take advantage of the **early bird registration until May 6, 2024**, we kindly request that one of the authors complete the registration by filling the following form https://forms.gle/HZ3Je3G9tJM7nJy4A

and join the whatsapp community for regular updates https://chat.whatsapp.com/CAL3ILqW1c4JcDvS1j4oMY

Please ensure to include the **paper id in the subject of the mail** and send the following documents to conferenceidea@gmail.com.

Paper 1

Registration Proof



Transaction Successful
11:57 am on 11 May 2024

Paid to

SO STUDENT ACTIVITY COUNCIL OIST ₹3,000

Transfer Details

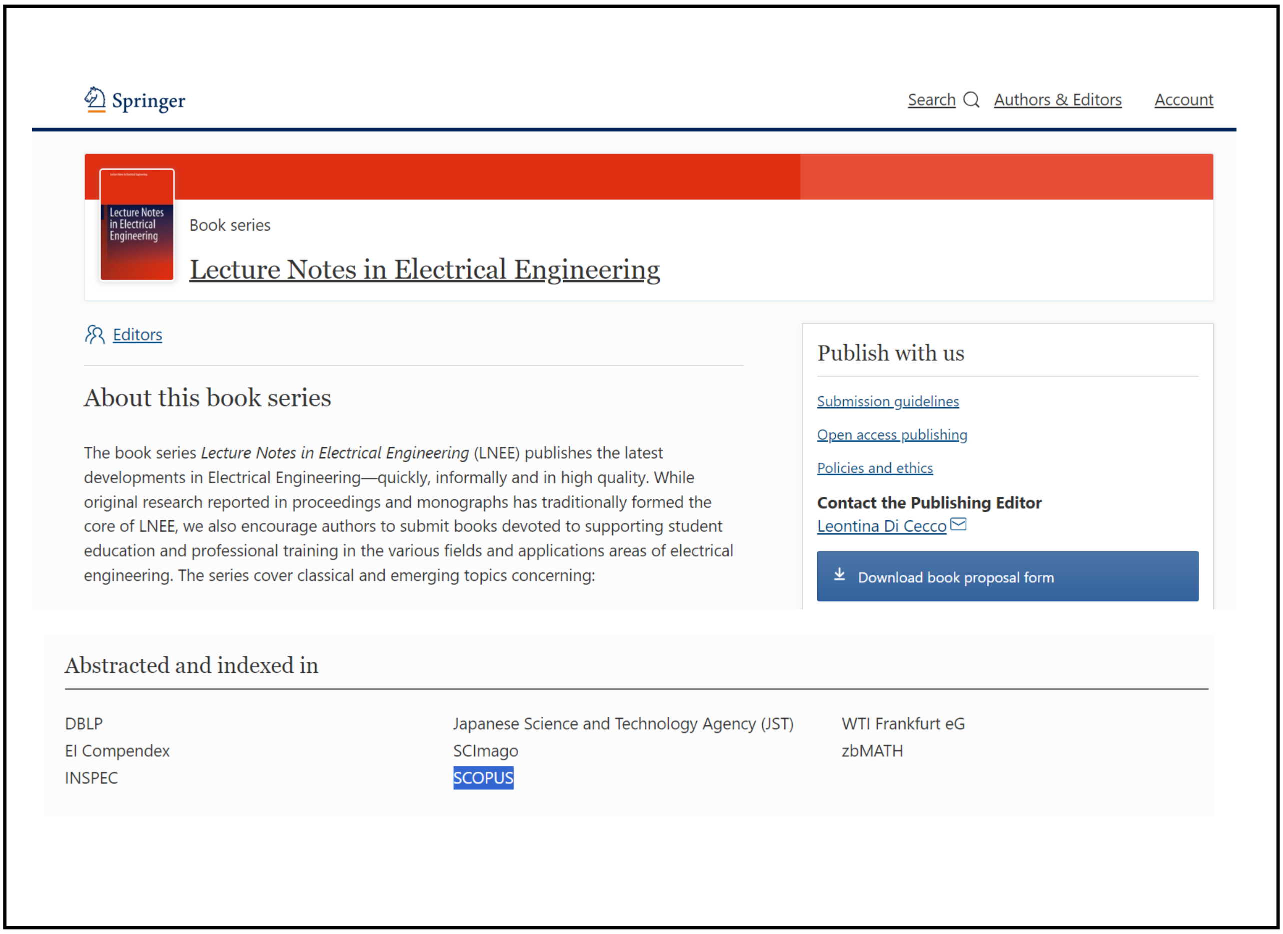
Transaction ID
T2405111150546045328358

Debited from
XXXXXX1185 ₹3,000
UTR: 413261474530

Powered by

Paper 1

Conference Indexing Details



Book series

Lecture Notes in Electrical Engineering

Editors

About this book series

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering—quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

Publish with us

Submission guidelines

Open access publishing

Policies and ethics

Contact the Publishing Editor
Leontina Di Cecco

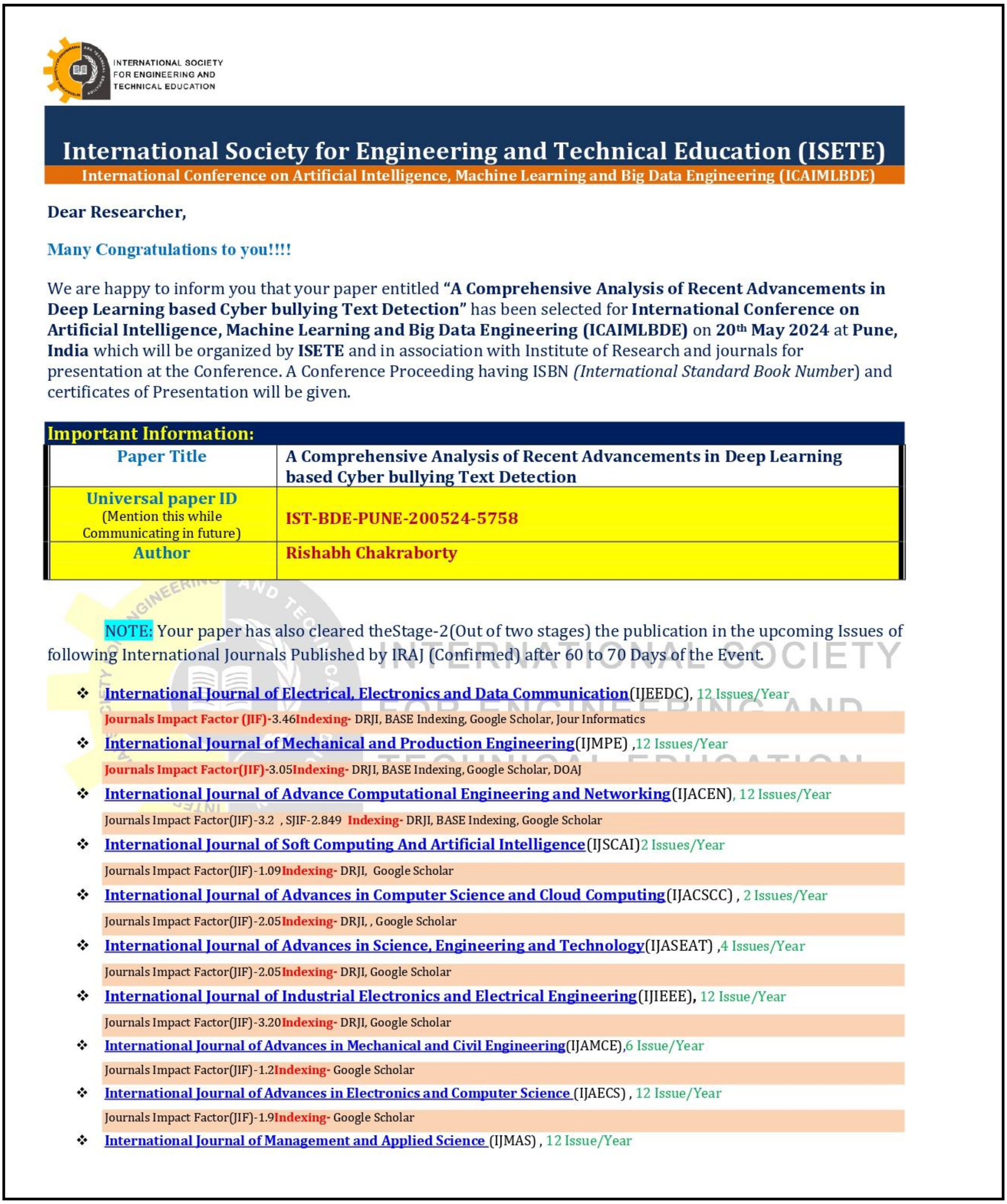
Download book proposal form

Abstracted and indexed in

DBLP
EI Compendex
INSPEC
Japanese Science and Technology Agency (JST)
SCImago
SCOPUS
WTI Frankfurt eG
zbMATH

Paper 2

Acceptance Notification



INTERNATIONAL SOCIETY
FOR ENGINEERING AND
TECHNICAL EDUCATION

International Society for Engineering and Technical Education (ISETE)

International Conference on Artificial Intelligence, Machine Learning and Big Data Engineering (ICAIMLBDE)

Dear Researcher,

Many Congratulations to you!!!!

We are happy to inform you that your paper entitled **"A Comprehensive Analysis of Recent Advancements in Deep Learning based Cyber bullying Text Detection"** has been selected for **International Conference on Artificial Intelligence, Machine Learning and Big Data Engineering (ICAIMLBDE)** on **20th May 2024** at **Pune, India** which will be organized by **ISETE** and in association with Institute of Research and journals for presentation at the Conference. A Conference Proceeding having ISBN *(International Standard Book Numbe*r) and certificates of Presentation will be given.

Important Information:

Paper Title	**A Comprehensive Analysis of Recent Advancements in Deep Learning based Cyber bullying Text Detection**
Universal paper ID (Mention this while Communicating in future)	**IST-BDE-PUNE-200524-5758**
Author	**Rishabh Chakraborty**

NOTE: Your paper has also cleared theStage-2(Out of two stages) the publication in the upcoming Issues of following International Journals Published by IRAJ (Confirmed) after 60 to 70 Days of the Event.

- **International Journal of Electrical, Electronics and Data Communication**(IJEEDC), 12 Issues/Year

  Journals Impact Factor (JIF)-3.46**Indexing**- DRJI, BASE Indexing, Google Scholar, Jour Informatics
- **International Journal of Mechanical and Production Engineering**(IJMPE) ,12 Issues/Year

  Journals Impact Factor(JIF)-3.05**Indexing**- DRJI, BASE Indexing, Google Scholar, DOAJ
- **International Journal of Advance Computational Engineering and Networking**(IJACEN), 12 Issues/Year

  Journals Impact Factor(JIF)-3.2 , SJIF-2.849 **Indexing**- DRJI, BASE Indexing, Google Scholar
- **International Journal of Soft Computing And Artificial Intelligence**(IJSCAI)2 Issues/Year

  Journals Impact Factor(JIF)-1.09**Indexing**- DRJI, Google Scholar
- **International Journal of Advances in Computer Science and Cloud Computing**(IJACSCC) , 2 Issues/Year

  Journals Impact Factor(JIF)-2.05**Indexing**- DRJI, , Google Scholar
- **International Journal of Advances in Science, Engineering and Technology**(IJASEAT) ,4 Issues/Year

  Journals Impact Factor(JIF)-2.05**Indexing**- DRJI, Google Scholar
- **International Journal of Industrial Electronics and Electrical Engineering**(IJIEEE), 12 Issue/Year

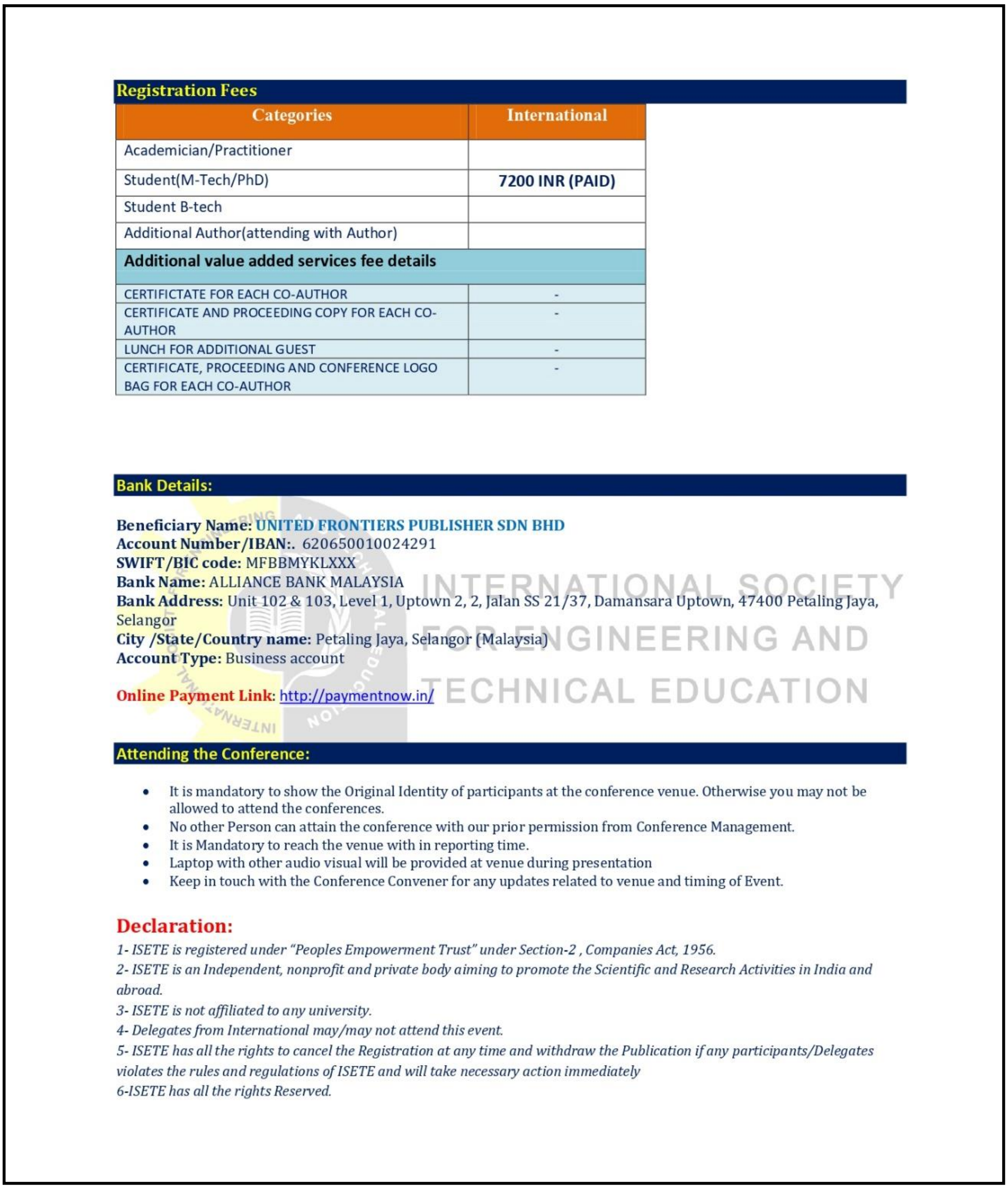
  Journals Impact Factor(JIF)-3.20**Indexing**- DRJI, Google Scholar
- **International Journal of Advances in Mechanical and Civil Engineering**(IJAMCE),6 Issue/Year

  Journals Impact Factor(JIF)-1.2**Indexing**- Google Scholar
- **International Journal of Advances in Electronics and Computer Science** (IJAECS) , 12 Issue/Year

  Journals Impact Factor(JIF)-1.9**Indexing**- Google Scholar
- **International Journal of Management and Applied Science** (IJMAS) , 12 Issue/Year

Paper 2

Registration Proof



Registration Fees

Categories	International
Academician/Practitioner	
Student(M-Tech/PhD)	**7200 INR (PAID)**
Student B-tech	
Additional Author(attending with Author)	
Additional value added services fee details	
CERTIFICTATE FOR EACH CO-AUTHOR	-
CERTIFICATE AND PROCEEDING COPY FOR EACH CO-AUTHOR	-
LUNCH FOR ADDITIONAL GUEST	-
CERTIFICATE, PROCEEDING AND CONFERENCE LOGO BAG FOR EACH CO-AUTHOR	-

Bank Details:

Beneficiary Name: UNITED FRONTIERS PUBLISHER SDN BHD
Account Number/IBAN:. 620650010024291
SWIFT/BIC code: MFBBMYKLXXX
Bank Name: ALLIANCE BANK MALAYSIA
Bank Address: Unit 102 & 103, Level 1, Uptown 2, 2, Jalan SS 21/37, Damansara Uptown, 47400 Petaling Jaya, Selangor
City /State/Country name: Petaling Jaya, Selangor (Malaysia)
Account Type: Business account

Online Payment Link: http://paymentnow.in/

Attending the Conference:

- It is mandatory to show the Original Identity of participants at the conference venue. Otherwise you may not be allowed to attend the conferences.
- No other Person can attain the conference with our prior permission from Conference Management.
- It is Mandatory to reach the venue with in reporting time.
- Laptop with other audio visual will be provided at venue during presentation
- Keep in touch with the Conference Convener for any updates related to venue and timing of Event.

Declaration:

1- ISETE is registered under "Peoples Empowerment Trust" under Section-2 , Companies Act, 1956.

2- ISETE is an Independent, nonprofit and private body aiming to promote the Scientific and Research Activities in India and abroad.

3- ISETE is not affiliated to any university.

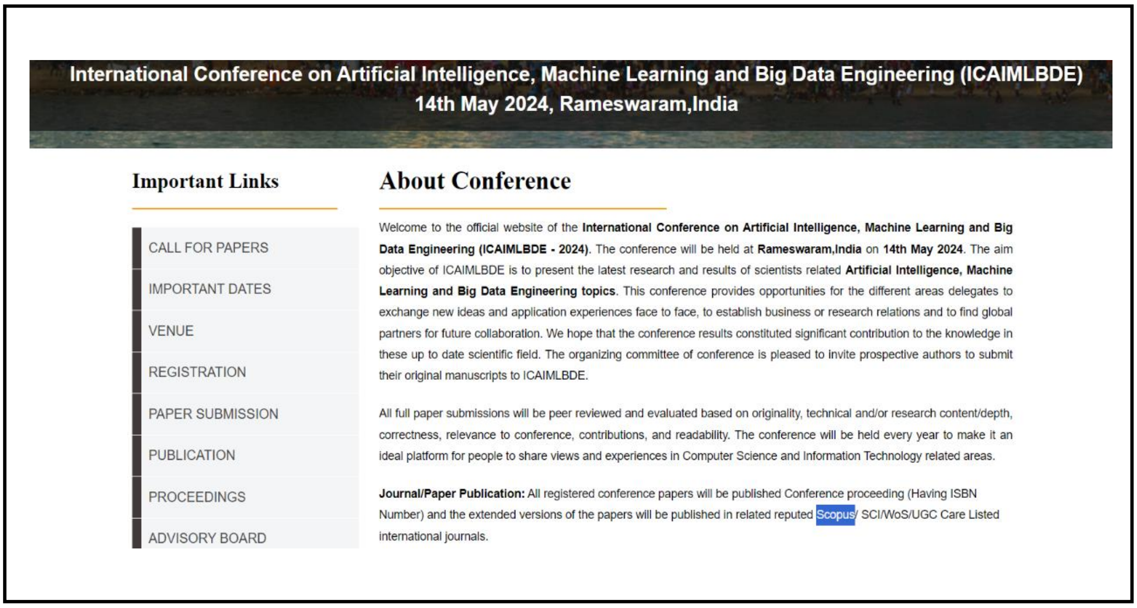
4- Delegates from International may/may not attend this event.

5- ISETE has all the rights to cancel the Registration at any time and withdraw the Publication if any participants/Delegates violates the rules and regulations of ISETE and will take necessary action immediately

6-ISETE has all the rights Reserved.

Paper 2

Conference Indexing Details



International Conference on Artificial Intelligence, Machine Learning and Big Data Engineering (ICAIMLBDE)
14th May 2024, Rameswaram,India

About Conference

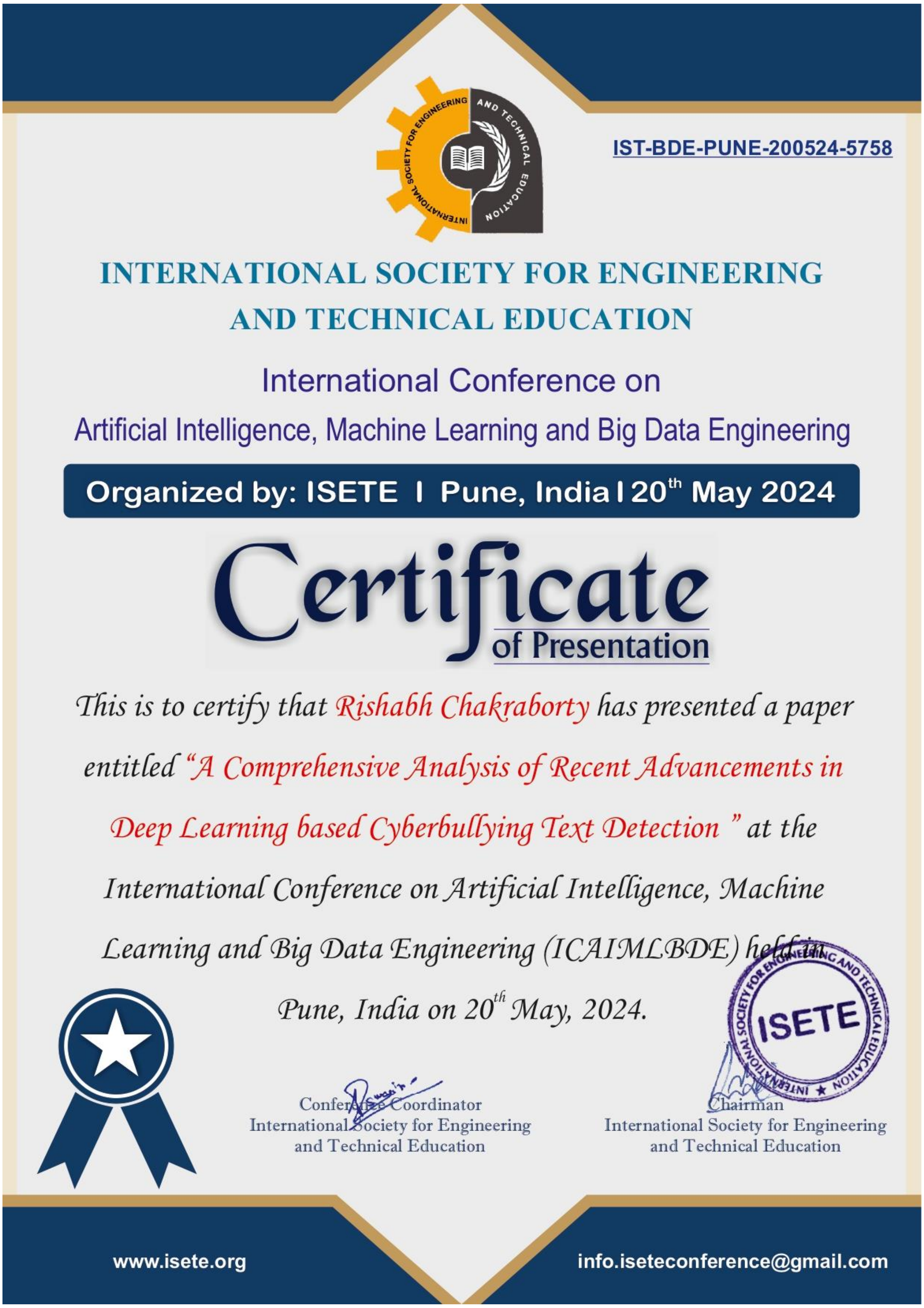
Welcome to the official website of the **International Conference on Artificial Intelligence, Machine Learning and Big Data Engineering (ICAIMLBDE - 2024)**. The conference will be held at **Rameswaram,India** on **14th May 2024**. The aim objective of ICAIMLBDE is to present the latest research and results of scientists related **Artificial Intelligence, Machine Learning and Big Data Engineering topics**. This conference provides opportunities for the different areas delegates to exchange new ideas and application experiences face to face, to establish business or research relations and to find global partners for future collaboration. We hope that the conference results constituted significant contribution to the knowledge in these up to date scientific field. The organizing committee of conference is pleased to invite prospective authors to submit their original manuscripts to ICAIMLBDE.

All full paper submissions will be peer reviewed and evaluated based on originality, technical and/or research content/depth, correctness, relevance to conference, contributions, and readability. The conference will be held every year to make it an ideal platform for people to share views and experiences in Computer Science and Information Technology related areas.

Journal/Paper Publication: All registered conference papers will be published Conference proceeding (Having ISBN Number) and the extended versions of the papers will be published in related reputed Scopus/ SCI/WoS/UGC Care Listed international journals.

Paper 2

Certificate



IST-BDE-PUNE-200524-5758

INTERNATIONAL SOCIETY FOR ENGINEERING AND TECHNICAL EDUCATION

International Conference on

Artificial Intelligence, Machine Learning and Big Data Engineering

Organized by: ISETE I Pune, India I 20th May 2024

Certificate

of Presentation

This is to certify that Rishabh Chakraborty has presented a paper entitled "A Comprehensive Analysis of Recent Advancements in Deep Learning based Cyberbullying Text Detection " at the International Conference on Artificial Intelligence, Machine Learning and Big Data Engineering (ICAIMLBDE) held in Pune, India on 20th May, 2024.

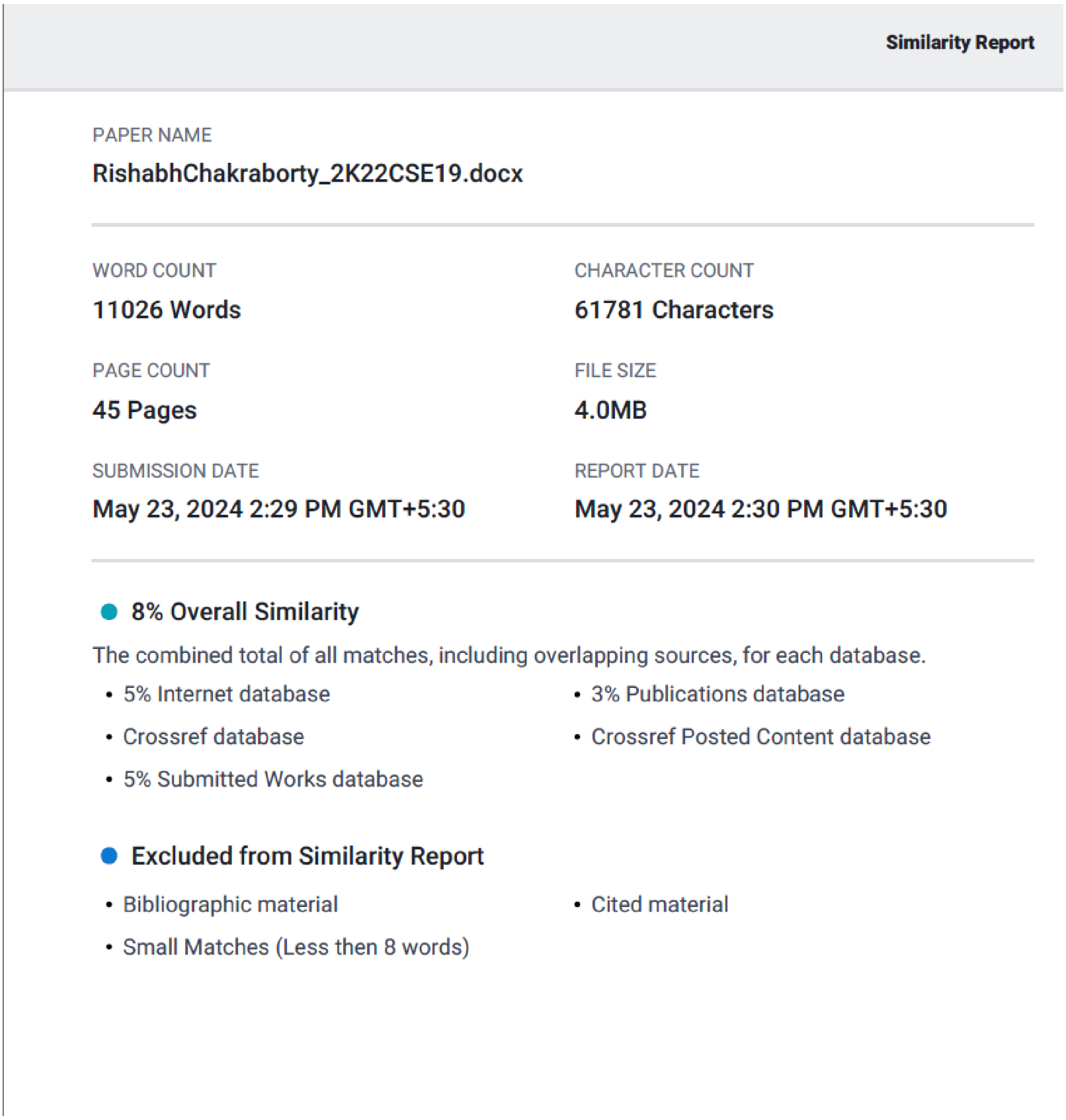
INTERNATIONAL SOCIETY FOR ENGINEERING AND TECHNICAL EDUCATION ★ ISETE

Conference Coordinator
International Society for Engineering
and Technical Education

Chairman
International Society for Engineering
and Technical Education

www.isete.org

info.iseteconference@gmail.com



Similarity Report

PAPER NAME

RishabhChakraborty_2K22CSE19.docx

WORD COUNT	CHARACTER COUNT
11026 Words	61781 Characters
PAGE COUNT	FILE SIZE
45 Pages	4.0MB
SUBMISSION DATE	REPORT DATE
May 23, 2024 2:29 PM GMT+5:30	May 23, 2024 2:30 PM GMT+5:30

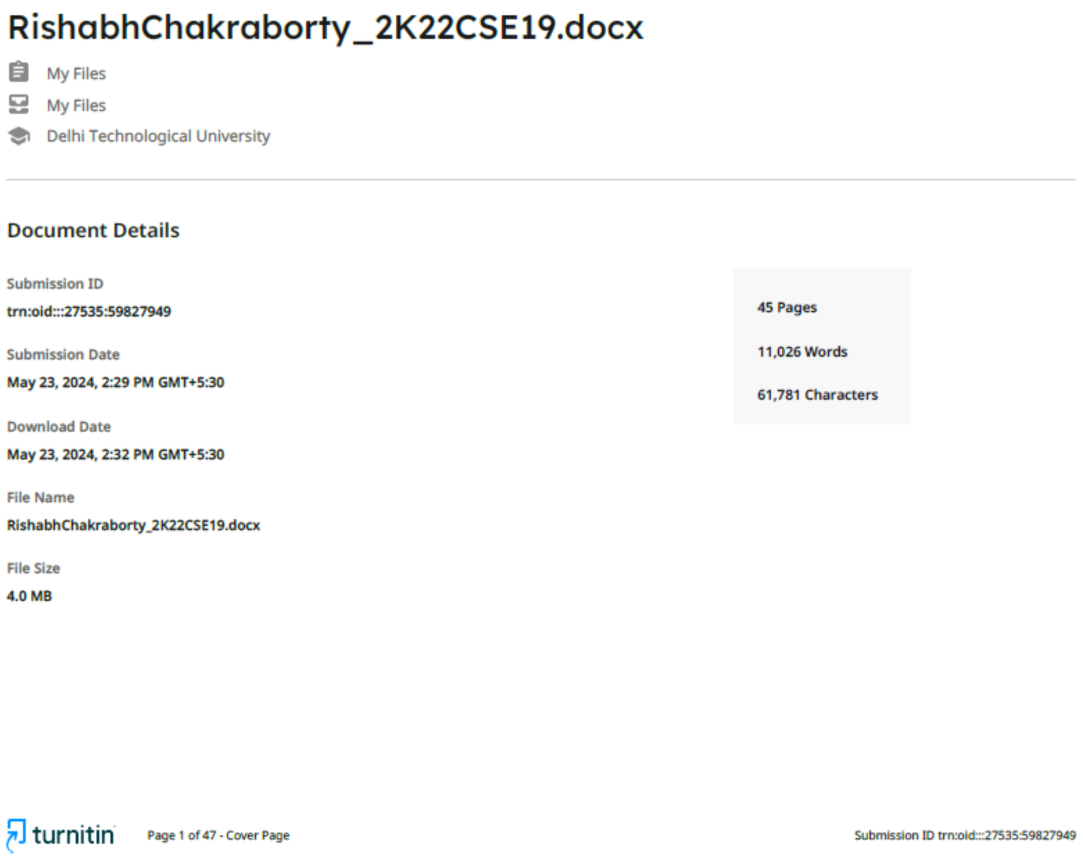
● 8% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 5% Internet database
- Crossref database
- 5% Submitted Works database
- 3% Publications database
- Crossref Posted Content database

● Excluded from Similarity Report

- Bibliographic material
- Small Matches (Less then 8 words)
- Cited material



RishabhChakraborty_2K22CSE19.docx

My Files

My Files

Delhi Technological University

Document Details

Submission ID

trn:oid:::27535:59827949

Submission Date

May 23, 2024, 2:29 PM GMT+5:30

Download Date

May 23, 2024, 2:32 PM GMT+5:30

File Name

RishabhChakraborty_2K22CSE19.docx

File Size

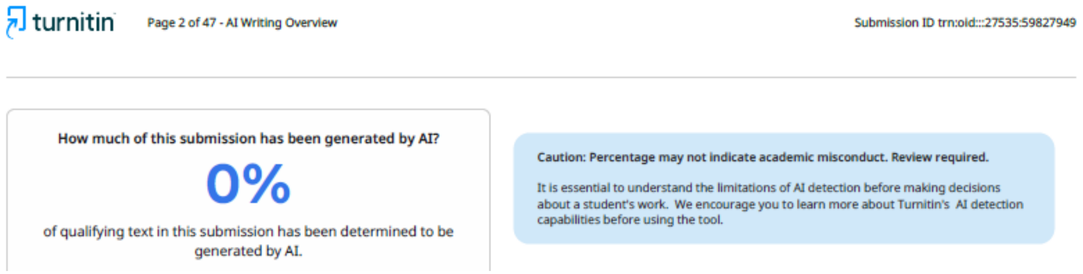
4.0 MB

45 Pages

11,026 Words

61,781 Characters

turnitin Page 1 of 47 - Cover Page Submission ID trn:oid:::27535:59827949



turnitin Page 2 of 47 - AI Writing Overview Submission ID trn:oid:::27535:59827949

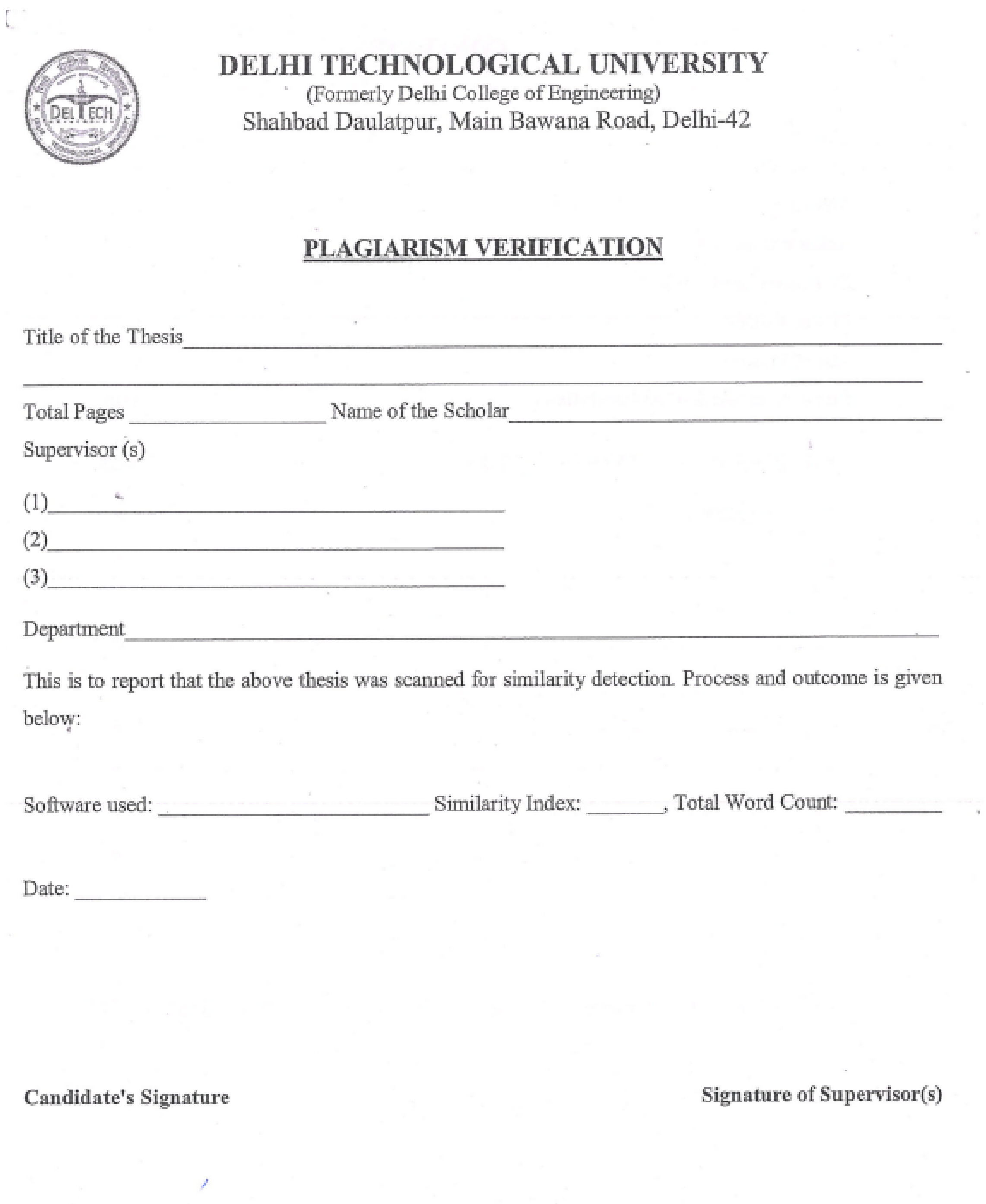
How much of this submission has been generated by AI?

0%

of qualifying text in this submission has been determined to be generated by AI.

Caution: Percentage may not indicate academic misconduct. Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.



DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis____

____

Total Pages _______________ Name of the Scholar___________________________________

Supervisor (s)

(1)____________________________________

(2)____________________________________

(3)____________________________________

Department____

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: ______________________ Similarity Index: _______, Total Word Count: _________

Date: ___________

Candidate's Signature **Signature of Supervisor(s)**

BRIEF PROFILE

I am Rishabh Chakraborty, pursuing my MTech in Computer Science and Engineering from Delhi Technological University. Currently, I am in the final semester of my degree and I scored 8.65 CGPA in the first three semesters of my MTech. I have also worked as a Placement Coordinator in TnP Department DTU from May 2023 onwards.

I completed my BTech in from VIT University, Vellore in 2020. After that, I worked in GeeksforGeeks as a Technical Content Writer and SME where I have published several articles and mentored a few students related to my field.

My area of interest is data science and data analytics. I learned numerous skills related to this area such as MS Excel. Power BI, SQL, Python, Snowflake, and so on. I am also very good at coding and solved more than 300 questions in GFG and LeetCode.