

# **Hate Speech Detection from Social Media Using Word Embeddings and Graph Neural Networks**

A MAJOR PROJECT REPORT II  
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE AWARD OF THE DEGREE  
OF

## **MASTER OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING**

Submitted by  
**Mr. ANURAG SHARMA**  
**Roll No-2K22/CSE/06**

Under the Supervision of  
**Dr. MINNI JAIN**  
**Assistant Professor**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**  
**DELHI TECHNOLOGICAL UNIVERSITY**  
(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-110042

MAY, 2024



# DELHI TECHNOLOGICAL UNIVERSITY

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

## CANDIDATE DECLARATION

I, **ANURAG SHARMA**, Roll no 2K22/CSE/06 student of M.Tech (Computer Science and Engineering), hereby declare that the Major Project Report II titled “**Hate Speech Detection from Social Media using Word Embeddings and Graph Neural Networks**” which is submitted by me to the Department of Computer Science and Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi

**ANURAG SHARMA**

Date:



# DELHI TECHNOLOGICAL UNIVERSITY

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

## CERTIFICATE

I hereby certify that the Major Project Report II titled “**Hate Speech Detection from Social Media using Word Embeddings and Graph Neural Networks**” which is submitted by **ANURAG SHARMA**, Roll No 2K22/CSE/06, Department of Computer Science and Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Date:

**Dr. MINNI JAIN**

(SUPERVISOR)

Assistant Professor,  
Department of CSE,  
DTU, Delhi

## **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 thesis.

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 report.

**Anurag Sharma**  
**(2K22/CSE/06)**

## **ABSTRACT**

The speedy expansion of online social networks has transformed the medium of communication and information promulgation, facilitating unprecedented levels of connectivity worldwide. OSN has increased social outreach as well and it also highlights any issues faster than conventional systems. However, this digital revolution has also provided a fascinating field for the procreation of hate speech, posing significant threats to both individual well-being and societal cohesion. In response to this pressing issue, researchers have vigorously pursued various methodologies aimed at identifying and detecting hate speech in OSN. Among these methodologies, deep learning techniques have emerged as particularly promising solutions as it provides more accurate results, leveraging their capacity to analyze vast amounts of textual data and extract meaningful patterns.

This major report undertakes a complete comparative analysis of different deep learning approaches for HSD, focusing intently on evaluating their performance using performance metrics. By analyzing each method on diverse datasets including Davidson-ICWSM, Waseem-EMNLP, Waseem-NAACL, and VLSP, we systematically evaluate the efficiency of multiple deep learning methods.

Convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory networks (LSTMs), transformers, graph convolutional networks (GCNs), and ensemble learning methodologies undergo stringent scrutiny in this study. Our analysis uncovers refinement insights into the strengths and limitations of each approach in the context of HSD. Among these, LR shows good results, but LSTM and bi-LSTM modeling have demonstrated exceptional performances, even though facing challenges such as handling multilingual datasets and classification issues. Additionally, BERT-based models show outstanding results in detecting derogatory language and slang across diverse linguistic landscapes.

Moreover, the introduction of graph convolutional network (GCN) models presents a promising approach for enhancing HSD capabilities. By capitalizing on the inherent structural relationships within online social networks, GCNs display notable potential in capturing complex patterns of HS propagation.

In conclusion, this major report serves as a valuable compass for advancing the field of HSD, tracing a course toward the development of more robust and effective strategies to uphold a safer and more inclusive digital environment.

**Keywords:** Hate Speech detection, BERT, Graph Convolutional Network, Bi-LSTM, CNN, RNN.

## **LIST OF PUBLICATIONS**

1. Anurag Sharma “Comparative Analysis of Deep Learning Techniques for Hate Speech Detection”, Accepted at the “**2<sup>nd</sup> International Conference on Machine Intelligence for Research and Innovations-2024 (MAITRI)**” at National Institute of Technology, Srinagar.

Paper id: 224

Indexed by Scopus.

2. Anurag Sharma “Advancements in Hate Speech Detection: A Comprehensive Review of Graph Convolutional Network (GCN) Models”, Presented at “**International Conference on Artificial Intelligence, Machine Learning and Big Data Engineering (ICAIMLBDE) organized by ISETE**” on 05<sup>th</sup> May 2024 at Hyderabad, India. [Presented]

Paper id: IST-BDE-HDBD-050524-5541

Indexed by Scopus.

## **TABLE OF CONTENTS**

|   |            |
|---|------------|
| <b>Candidate's Declaration .....</b>              | <b>i</b>   |
| <b>Certificate .....</b>                          | <b>ii</b>  |
| <b>Acknowledgement .....</b>                      | <b>iii</b> |
| <b>Abstract .....</b>                             | <b>iv</b>  |
| <b>List of Publications .....</b>                 | <b>vi</b>  |
| <b>List of Tables .....</b>                       | <b>ix</b>  |
| <b>List of Figures .....</b>                      | <b>x</b>   |
| <b>List of Abbreviations .....</b>                | <b>xi</b>  |
| <b>CHAPTER 1 INTRODUCTION .....</b>               | <b>01</b>  |
| 1.1 A brief overview .....                        | 01         |
| 1.2 Motivation .....                              | 02         |
| 1.3 Problem Statement .....                       | 03         |
| 1.4 Working .....                                 | 04         |
| 1.5 Report Outline .....                          | 04         |
| <b>CHAPTER 2 BACKGROUND .....</b>                 | <b>06</b>  |
| 2.1 Hate Speech .....                             | 06         |
| 2.2 Hate Speech Detection .....                   | 06         |
| 2.3 Preprocessing .....                           | 08         |
| 2.4 Datasets .....                                | 09         |
| <b>CHAPTER 3 LITERATURE REVIEW .....</b>          | <b>11</b>  |
| 3.1 Machine Learning and Deep Learning .....      | 11         |
| 3.1.1 Machine Learning .....                      | 11         |
| 3.1.2 Deep Learning .....                         | 12         |
| 3.2 Word Embeddings .....                         | 13         |
| 3.2.1 Word2Vec .....                              | 13         |
| 3.2.2 Glove Embedding .....                       | 13         |
| 3.3 Single and Hybrid ML Algorithms for HSD ..... | 13         |



|   |           |
|---|-----------|
| 3.4 Different Techniques based on Deep Learning for HSD ..... | 15        |
| 3.4.1 CNN .....   | 15        |
| 3.4.2 RNN .....   | 15        |
| 3.4.3 LSTM .....  | 16        |
| 3.4.4 Transformers .....                                      | 17        |
| 3.4.5 GCN .....   | 18        |
| 3.4.6 Ensemble Learning .....                                 | 19        |
| 3.4.7 RBM .....   | 19        |
| 3.4.8 AE .....  | 20        |
| 3.4.9 Bi-LSTM .....   | 21        |
| <b>CHAPTER 4 LITERATURE SURVEY .....</b>                      | <b>22</b> |
| 4.1 Literature Survey of Deep Learning Models .....           | 22        |
| 4.2 Literature Survey of Various GCN Models .....             | 23        |
| <b>CHAPTER 5 MODEL ANALYSIS .....</b>                         | <b>26</b> |
| <b>CHAPTER 6 RESULT AND DISCUSSION .....</b>                  | <b>28</b> |
| 6.1 Performance Metrics .....                                 | 28        |
| 6.2 Comparison of various Deep Learning Techniques .....      | 28        |
| 6.3 Comparative Analysis of Various GCN Models .....          | 31        |
| <b>CHAPTER 7 CONCLUSION .....</b>                             | <b>33</b> |
| <b>REFERENCES .....</b>                                       | <b>34</b> |

## **LIST OF TABLES**

|           |  |    |
|-----------|--|----|
| Table 2.1 | Converting data into pre-processed form .....    | 09 |
| Table 2.2 | Datasets description .....                       | 10 |
| Table 3.1 | Evaluating methods on different parameters ..... | 14 |
| Table 6.1 | Showing various parameters in percentage .....   | 29 |
| Table 6.2 | Performance Evaluation of GCN Models .....       | 32 |

## **LIST OF FIGURES**

|              |  |    |
|--------------|--|----|
| Figure 2.1   | Showing steps of HSD .....                     | 08 |
| Figure 3.1   | CNN .....                                      | 15 |
| Figure 3.2   | RNN .....                                      | 15 |
| Figure 3.3   | Architecture of LSTM .....                     | 15 |
| Figure 3.4.1 | MLM (Masked Language Model) .....              | 17 |
| Figure 3.4.2 | NSP (Next Sequence Prediction) .....           | 18 |
| Figure 3.5   | GCN architecture .....                         | 19 |
| Figure 3.6   | RBM architecture .....                         | 20 |
| Figure 3.7   | Simple AE architecture .....                   | 20 |
| Figure 3.8   | Architecture of Bi-LSTM .....                  | 21 |
| Figure 6.1   | Accuracy chart of DL techniques .....          | 29 |
| Figure 6.2   | Precision Chart of various DL techniques ..... | 30 |
| Figure 6.3   | Recall chart of DL techniques .....            | 30 |
| Figure 6.4   | Graph showing F1 Score of various models ..... | 31 |
| Figure 6.5   | Performance scores of various GCN models ..... | 32 |

## **LIST OF ABBREVIATIONS**

|      |                                |
|------|--------------------------------|
| HS   | Hate Speech                    |
| HSD  | Hate Speech Detection          |
| OSN  | Online Social Network          |
| DL   | Deep Learning                  |
| ML   | Machine Learning               |
| GNN  | Graph Neural Network           |
| CNN  | Convolutional Neural Network   |
| RNN  | Recurrent Neural Network       |
| GCN  | Graph Convolutional Network    |
| LSTM | Long Short-Term Memory         |
| DGC  | Dependency Graph Convolutional |
| NN   | Neural Network                 |
| SDG  | Syntactic Dependency Graph     |
| NLP  | Natural Language Processing    |
| L-R  | Left to Right                  |
| R-L  | Right to Left                  |
| SM   | Social Media                   |
| AI   | Artificial Intelligence        |
| DNN  | Deep Neural Network            |
| I/P  | Input                          |
| O/P  | Output                         |
| VS   | Vector Space                   |
| WWW  | World Wide Web                 |
| LBC  | Lexicon-based classification   |
| R&D  | Research and Development       |

# **CHAPTER 1**

## **INTRODUCTION**

### **1.1A Brief Overview**

In the contemporary digital landscape, people are connected to each other over an extensive network. Digital media platforms function as spaces for the expression of thoughts and the analysis of diverse themes. Although this connectivity enables individuals to exchange information and encourages enthusiasm, it has also led to the emergence of a more negative aspect. Certain individuals take advantage of this interconnection to spread vulgar and harmful content about others online, causing negative effects on the general public. Notable individuals, along with regular people, have been subjected to overt harassment, leading to loss of morality and a deterioration of thy mental health. The exponential expansion of online platforms has intensified the problem of hate speech, leading to an unfavourable atmosphere for the majority of users. The use of disrespectful language has the potential to intensify hostility between various ethnic, racial, linguistic, or local social groups, media outlets, or organizations. It is essential to promptly, effectively, and flexibly deal with hate speech, considering the impracticability of manually verifying extensive amounts of content. There are legal provisions, such as Section 153A, that allow for criminal charges to be filed against individuals who engage in contemptuous speech.

Any communication, expression, or phrase that targets and creates discrimination among the people or against certain minority groups based on qualities like gender, race, ethnicity, religion, sexual orientation, or other distinguishing traits is referred to as hate speech. It crosses the line into polite conversation and frequently uses harsh or disparaging language in an attempt to hurt, marginalize, or inspire violence. Hate speech is a major obstacle to the development of inclusive and tolerant society since it can appear in a variety of forms, such as written text, spoken communication, and internet platforms. Intending to expertly address and counter hate speech prevailing in the social media, it is necessary to have a thorough grasp of all of its forms as well as to adopt policies that will encourage civil and polite public discourse.

The requisite of the present situation is to affray this genuine issue and which has motivated numerous professionals and scholars to explore innovative methodologies steered at recognizing and dealing with HS in online content. Different supervised and unsupervised techniques, such as Bayesian models, BERT models, CNN, GCN, RNN, and SVM, have functioned to detect & palliate harmful material effectively. Amidst these efforts, Graph Convolutional Networks (GCNs) have turned up as a superior approach to understanding the intricate dynamics of HSD within OSN. GCNs offer a distinctive method for analysing the structural and semantic relationships inherent in social media data. This helps in capitalizing the underlying trends of the corpus. By capitalizing on the inherent graph structure of online interactions, GCNs facilitate the extraction of steep insights into the underlying patterns and mechanisms driving the spread of prevailing HS. Unlike conventional ML algorithms that operate on vectorized data representations, GCNs directly process graph-structured data by making a co-occurrence matrix, enabling a more nuanced examination of interconnected nodes and their relationships.

This report makes efforts to explore the efficacy of GCN in detecting HS within OSD. By harnessing the power of graph-based representations, the research aims to unravel the intricate web of HS propagation, identify influential nodes and communities, and develop robust models capable of accurately discerning HS from gracious content.

By a comprehensive evaluation of literature written in past, methodologies, and datasets, the thesis roots to elucidate the potential of GCNs as a tool for combating online hate speech. By conducting empirical studies and comparative analyses, we pitch attention on the strengths, limitations & practical implications of scouting GCNs in HSD tasks.

Ultimately, this research aspires to contribute to the working efforts to foster a hale, more inclusive environment (online) by leveraging cutting-edge techniques such as Graph Convolutional Networks to counter vicious effects of HS.

## **1.2 Motivation**

The motivation for HSD in OSD using graph neural network(s) stems from the prior want to acknowledge the rising tide of vicious content booming on OSD. Social media, while offering unprecedented connectivity and information enduring, primarily become a

breeding ground for HS, posing significant threats to individual well-being and societal cohesion.

Orthodox methods of HSD often plunge short in sufficiently annexing the complex dynamics & shade inherent in online interactions. GNNs tender a rising solution by leveraging the inherent graph structure of social media data. By modelling the link between users, posts, and interactions as a graph, GNNs enable a more holistic perceptive of the propagation patterns of HS within OSD.

The catalyst behind utilizing GNNs lies in their skill to hook both the structural & semantic information embedded within social media networks. Unlike traditional ML algorithms that compel on vectorized representations of data, GNNs directly process graph-structured data, allowing for a more nuanced analysis of interconnected nodes and their interdependencies.

Furthermore, by advancing the potential of graph convolutional layers, GNNs can efficaciously aggregate information from adjoining nodes, enabling them to spot subtle patterns of HS propagation that may not be apparent through orthodox methods.

In essence, the motivation for HSD using GNN pretense in the quest for more effective and comprehensive solutions to combat online HS, conclusively cherishing a safer and more comprehensive digital world for all users.

### **1.3 Problem Statement**

In response to the intensifying concerns surrounding HS and its pernicious effect on users' well-being in OSD, this thesis endeavors to address, the imperative demand for advanced computational methods for automatically detecting HS text on OSD. The study specifically pursues to probe the effectiveness of employing a graph neural network architecture and word embeddings in identifying hateful content within texts. Recognizing the injunction of old methods in seizing the nuanced, context-dependent nature of hate speech behaviors, the research aims to develop more sophisticated strategies.

By studying various neural network (NN) architectures, the study compares conventional NN approaches with hybrid-based techniques. Additionally, the thesis conducts a comprehensive review of previous methodologies, retaining a spectrum of DL techniques and hybrid models, to classify their performance using a confusion matrix. Through this investigation, the thesis endeavors to shed light on emerging concepts and methodologies utilized in addressing cyber hate.

Furthermore, the research includes a comparative analysis of different Graph Convolutional Network (GCN) models proposed by various researchers for detecting and addressing HS on OSD.

Depending on the problem statement following questions can be identified:

1. What are the various techniques used in HSD?
2. What are deep learning techniques?
3. What is the outcome of various GCN models?
4. What are the advancements in each model?
5. What is the comparative analysis of each model?

## **1.4 Working**

This research methodology involves accessing multiple databases such as "IEEE Explore," "Scopus," "ACM," "Science Direct," and "Kaggle" to gather relevant articles for investigating cyber hate speech. The focus is on identifying papers related to cyberbullying, cyber hate, and toxic speech within the sphere of NLP. To ensure the selection of recent & pertinent papers, a filtering tool was utilized to limit the search to the past seven years. The findings and advancements reported in these selected papers are thoroughly discussed in the preceding sections of this thesis.

## **1.5 Report Outline**

The sections within the report are as mentioned below:

### **Chapter 1- Introduction**

Introduce the problem statement and motivation for writing this report as well as described in brief the details about my work in the report.

### **Chapter 2- Background**

Described about the hate speech, challenges and techniques to tackle and detect it. Briefly discussed about the preprocessing and datasets used in the later section of the report.



### Chapter 3- Literature Review

Studied the detailed explanation, working of various ML and deep learning methods used for the HSD.

### Chapter 4 – Literature Survey

Studied the various research papers and summarise the key points of each paper related to the title of the report.

### Chapter 5- Model Analysis

Reviewed the working algorithms and implementation various GCN models used for the HSD.

### Chapter 6- Comparative Analysis and Discussion

Compared and analysed the various existing models and also visualised the performance scores of each model.

### Chapter 7- Conclusion

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

## **CHAPTER 2**

### **BACKGROUND**

The swift expansion of OSD has facilitated unprecedented global digital reach and interaction among individuals from diverse countries, cultures, and ethnic backgrounds. While this interconnectedness offers numerous advantages and positive outcomes, it has also paved the way for the proliferation of xenophobic, racial, and sexist remarks [1]. The anonymity, accessibility, and lack of accountability inherent in online platforms have emboldened individuals to express themselves in ways that would be inhibited in face-to-face interactions.

#### **2.1 Hate Speech**

HS means any setup of communication that denigrates, or libels individuals or groups on characteristics like race, ethnicity, gender, sexual orientation, or disablement, which has become a critical issue in synchronous society, particularly in online spaces. HS is nowadays creating a negative impact on society. It is also becoming a harmful tool to target someone's character. The exponential growth of OSD has furnished a breeding space for the diffusion of HS, posing significant threats to one's well-being, social harmony, and freedom of speech.

HS, portrayed by expressions of animosity or contemptuous language directed at specific groups, has assembled elevated attention from OSD like Facebook, which have emphasized their efforts to combat such harmful content [2]. Despite these efforts, platforms concede the trouble in adequately detecting and removing such material [3]. Similarly, companies like X are reassessing their policies to address offensive behavior and implement new steps to deal with hateful content, including issuing warnings, deleting harmful tweets, or suspending users [4] within a specific time frame. However, the humongous volume of data generated by the user creates a significant challenge for these platforms, despite their endless investment in resources and manpower [5].

#### **2.2 Hate Speech Detection**

The rise of OSD such as Facebook, X, and Instagram has reinforced the importance of detecting and treating HS in OSN. HSD in virtual communities aims to identify and alleviate harmful content. With the continuing expansion of the WWW and the increasing

occurrence of online engagement, users of social media, blogs, and forums often find themselves susceptible to harassment. This online criticism can have far-reaching consequences, impacting both the individuals targeted and the wider online community. Many strict rules and guidelines have been generated, but it hasn't controlled online criticism. Examples of such abuse include statements like "members of a specific caste are employed by locals" or "individuals from a specific religious group should be disintegrated. " These tweaked messages, directed at minority groups, are unfortunately all too common in OSN [6].

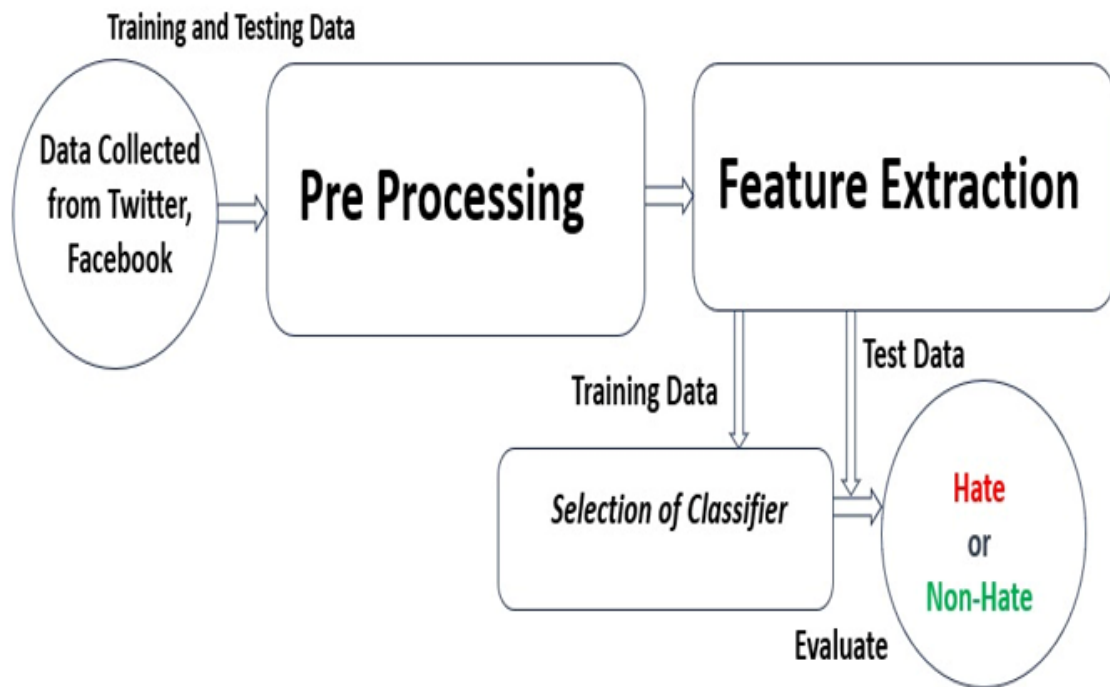
Recently, there has been a rapid faster movement on the internet advocating for the protection of minority groups [7]. This movement corresponds with AI for the Social Good (AI4SG) principles. Additionally, discussions connecting issues such as sexual harassment [8], sexual discrimination detection [9], cyberbullying, and trolling have been prevalent in academic papers [10]. Recent research has delved into the detection of suicidal phantasy as a means of presenting the real-world consequences of hatred on the internet [11]. This leads to an increase in suicidal cases across the globe. The earliest studies on HSD relied on bag-of-words (BOW) approaches [12][13], laying the basics for detailed research in this field. In distinctiveness to pattern-based methods, a study in 2012 introduced classifiers based on computers for identifying HS [14][15].

Orthodox ML methods, including logistic regression (LR), support vector machines (SVM), and decision trees (DT), have been commonly used for the HSD [12]. Moreover, non-linguistic features such as the author's gender or ethnicity can contribute to improving the classification accuracy of hate speech, although such data are often unavailable or unreliable on social media platforms [13]. New datasets classify the emotion more accurately than previous datasets. Sentiment analysis and polarity detection methods are also frequently used for HSD, given that hateful language is inherently a negative emotion. Therefore, messages conveying negative emotions are more likely to display hate compared to neutral or positive messages [16].

Additionally, external lexical tools have been employed in hate speech detection (HSD), drawing inspiration from sentiment analysis and affective computing [17]. For instance, [18] introduced a hate verb lexicon designed to identify verbs that endorse or advocate violent actions. However, the reliability of LBC largely hinges upon consistency of outer

resources. To mitigate this dependency, some initiatives have merged the benefits of machine learning (ML) with classification methods based on lexicons for hate speech detection [19].

Recently, there has been a notable shift towards employing neural patterns for hate speech detection. These models often leverage deep learning (DL) techniques, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory Networks (LSTMs) [20], which have demonstrated significant efficacy across various natural language processing (NLP) tasks..



**Figure 2.1** Showing steps of HSD

## 2.3 Preprocessing

Preprocessing for hate speech detection encompasses cleaning and standardizing text data, including tasks like removing special characters, the removing characters such as |: ,; &? ; portrayed. converting to lowercase, tokenizing, eliminating stop words, stemming, normalizing, handling emoticons/emojis, URLs/user mentions, contractions, profanity filtering, class balancing, and text vectorization. Table 1 shows the results after the preprocessing of imbalanced data. Additionally, normalizing hashtags in the data entails converting them into words; for example, "#hateblackpeople" is transformed into "hate black people".

**Table 2.1:** Converting data into pre-processed form

| Steps                     | Original   | Pre-processed  |
|---------------------------|--|--|
| Punctuation marks removal | Hello India testing words for 9 reviews @ Delhi University | Hello India testing words for 9 reviews Delhi University |
| Convert to lowercase      | Hello India testing words for 9 reviews Delhi University   | Hello india testing words for 9 reviews delhi university |
| Deletion of numerical     | Hello india testing words for 9 reviews delhi university   | Hello india testing words for reviews delhi university   |
| Correction of spelling    | Hello india testing words for reviews delhi university     | Hello india testing words for reviews None university    |
| Singularization           | Hello india testing words for reviews None university      | Hello india testing word for review None university      |
| Converting in Base form   | Hello india testing word for review None university        | Hello india test word for review None university         |
| Removal of Stop-words     | Hello india test word for review None university           | Hello india test word review university                  |

## 2.4 Datasets

In HSD, a dataset comprises annotated text data indicating whether each bit of text holds HS, offensive language, or is non-offensive. These datasets serve as training data for ML models to learn patterns and features associated with HS.

The datasets utilized in HSD studies are pivotal to train & assess the performance of machine learning models. Below are descriptions of several commonly employed datasets in hate speech detection research:

1. **Davidson-ICWSM Dataset** [23]: These data encompass tweets which are categorized as offensive language, or not offensive content and also, HS. It provides a wide-ranging collection of tweets taken from Twitter, with annotations given by human annotators.
2. **Waseem-EMNLP Dataset** [21]: The Waseem-EMNLP dataset includes tweets marked for hate speech in English language. It has annotations for various types of HS, such as sexism, racism, and homophobia.

3. **Waseem-NAACL Dataset** [22]: Similar to Waseem-EMNLP dataset, this data contains tweets specifically marked for hate speech. The main focus is on identifying hate speeches aimed at women and religious groups.
4. **VLSP Dataset** [24]: The Vietnamese Social Media Hate Speech (VLSP) dataset is made up of posts taken from various social media platforms in Vietnamese. This dataset is marked for different categories of HS and offensive content commonly found in Vietnamese internet communities.

**Table 2.2:** Datasets description

| Dataset        | Tags                              | Count of non-hateful instances | Count of Hateful instances | Total instances |
|----------------|-----------------------------------|--------------------------------|----------------------------|-----------------|
| Waseem-EMNLP   | Sexism, Racism, neither, and both | 5850                           | 1059                       | 6909            |
| Waseem-NAACL   | Racist, Sexist, and neither       | 11501                          | 5406                       | 16910           |
| Davidson-ICWSM | Offensive, hate speech, neither   | 4163                           | 20620                      | 24783           |
| VLSP-HSD       | Clean, hate and offensive         | 18614                          | 1731                       | 20345           |

## **CHAPTER 3**

### **LITERATURE REVIEW**

#### **3.1 Machine Learning and Deep Learning**

OSNs tend as arenas for the personage to openly eloquent his/her thoughts and opinions without agitation of being subjected to hostile actions. HS, which refers to communication that expresses furious hatred or enforces violence based on certain qualities, presents a revelatory danger to this principle. It includes misinformation, the approbation of violence, and prejudice against persons or groups based on different factors. The use of inflammatory language in OSD not only disrupts an individual's serenity but also contributes to the development of mental health problems and affects the peace and harmony of society. It is incumbent to tackle this issue, as unattended hate speech has the potential to grow into grave offenses, physical aggression, and disputes.

Explicit harassment has a consequential impact on the internet community, hurting both regular folks and public figures. HS detection tools are crucial for acknowledging this crucial problem on OSD, guaranteeing an applauded social atmosphere. Maintaining a balance between HS identification and the protection of freedom of expression is very crucial. Distinguished instances of HS encircle profanity, humiliating women, remarks/comments regarding physical attributes (sex, color), comparisons, sweeping statements, and mockery of events.

Top SM platforms have agreed to abide by a code of conduct and have promised to swiftly review and delete/censor illegal HS within 24 hours. Machine-based algorithms that are capable of identifying HS are necessary since SM datasets are large and dynamic. Numerous DL and ML models have shown promising results in this field.

##### **3.1.1 Machine Learning**

Within the study of artificial intelligence, ML focuses on developing complex algorithms that can understand underlying patterns, and trends and make rulings/predictions without the need for definitive programming. It uses data analysis and visualization to identify complex patterns and linkages, enhancing research across a range of fields.

## **Supervised Learning**

Algorithms are taught on a labeled dataset in a particular ML technique called “supervised learning”. In other words, the algorithm learns the mapping between inputs and associated outputs by being fed input-output pairs during the training phase. This approach works especially well for tasks like regression and classification because it allows the computer to generalize its learning to predict outcomes on fresh, unobserved data

## **Unsupervised Learning**

This learning engages unlabeled datasets for the training of algorithms. In this case, the system explores the underlying structures or patterns in the data without explicit direction. Clustering, in which the algorithm locates natural groupings in the datasets, and dimensionality reduction is applied, which helps to find underlying links without specified labels. Unsupervised learning is most often used for tasks like anomaly detection and data exploration, and also it aids in the discovery of intrinsic data features.

### **3.1.2 Deep Learning**

DL is a specialistic field within ML that harnesses deep NN, which consists of collective layers, to gear complex problems with bizarre efficacy. DL is an algorithm that sovereignly acquires hierarchical representations of data by processing input through interconnected layers of nodes or artificial neurons. DL actually presents the true sense of AI. The hierarchical structure allows DL models to effectually attain complex patterns and features in the data, which the human brain needs plenty of time. This makes them highly productive in jobs like picture and speech recognition, NLP, and other complex pattern identification challenges.

## **Supervised Deep Learning**

It is a distinctive path in the domain of DL, wherein the algorithm endures training using a dataset that has been labeled. Like orthodox supervised learning, the DL model is presented with input-output pairs amid the phase of training. This phase is the most important aspect of DL. The DNN cultivates the feature to establish I/P data with fitted O/P labels by iteratively advancing its internal parameters along with hyperparameters using the backpropagation algorithm. This technique is highly effective in tasks that require classification, regression, and prediction.



## **3.2 Word Embeddings**

Word embeddings are the method of converting/assigning each word with particular numerical values in a continuous VS for NLP tasks. They encode semantic relationships and contextual meanings, positioning words having similarity meanings near together in the VS. This departure from treating words as discrete symbols has glorified NLP applications by seeking subtle linguistic features.

Popular word embedding models are discussed in the sub-sections. These models got training on larger datasets to learn distributed word representation(s) based on co-occurrence patterns. Aside from encoding syntactic and semantic relationships, word embeddings enable mathematical operations such as analogies.

### **3.2.1 Word2Vec**

Word2Vec is a widely used word embedding technique. It predicts a word's surrounding context or neighboring words within a corpus. Word2Vec offers two model architectures: Skip-Gram and Continuous Bag of Words (CBOW). Skip-Gram foretells context words given a target word, while CBOW predicts the target word based on its context. After training on large datasets, Word2Vec prompts dense vector representations for words, capturing semantic similarities and relationships. These embeddings remediate the effectiveness of downstream NLP applications and have demonstrated efficacy across various tasks.

### **3.2.2 Glove Embedding**

Glove is a prominent word embedding method that influences a corpus's overall statistical information. It creates word embeddings by analyzing the global (co-occurrence data) of words in the corpus. By factorizing the word co-occurrence matrix, Glove embeddings encode semantic relationship(s) among words. They excel in jobs mainly word analogies and semantic similarity due to their ability to seek complicated syntactic & semantic traces within available data. The training process categorizes seeking general word distributional patterns, resulting in embeddings that effectively represent complex linguistic relationships within the dataset.

## **3.3 Single and Hybrid ML Algorithms for HSD**

HSD is a rigorous task that has prompted the exploration of many ML techniques, both single and hybrid, to adequately address the cost and complexity of identifying and mitigating hateful content online [25].

Here, we will discuss few commonly employed approaches:

### **Single Methods**

This methodology serves as a detector for categorizing hate speech within Twitter data through the utilization of a single machine learning (ML) classification approach. Additionally, ML is employed for both extracting & pre-processing X data across various volumes. ML is widely recognized as an algorithmic and statistical technique for addressing diverse problem sets.

### **Hybrid Methods**

This approach amalgamates different ML techniques to improve the effectiveness of conventional human method(s). It is perceived as an advancement over individual machine learning methods for achieving enhanced results in identifying hate speech on Twitter. Hybrid methodologies are deemed particularly robust, and they are seen as better equipped to handle the substantial abundance of metadata generated on OSN. These hybrid models are characterized by greater computational efficiency and are known to deliver superior performance compared to their singular counterparts.

**Table 3.1** Evaluating methods on different parameters

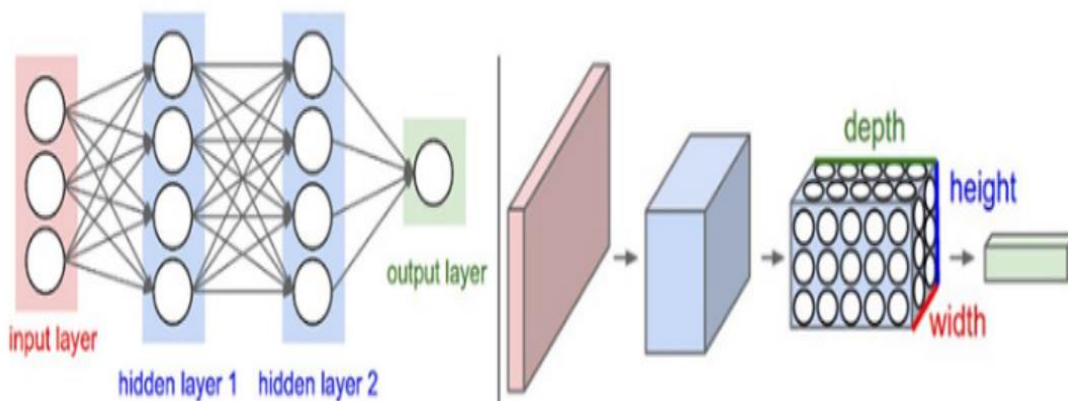
| <b>Methods</b> | <b>Benefits</b>                           | <b>Drawbacks</b>   | <b>Approaches</b>  |
|----------------|---|--|--|
| Single Method  | Stability,<br>Adaptable and<br>Extensible | Low Precision,<br>fragmentation<br>problem,<br>imbalance data<br>performance | 1.ANN (RNN,<br>CNN, MLP)<br>2.DL (LSTM,<br>CNN-1D)<br>3.GA(GP)<br>4. LR<br>5.Kernel Methods<br>(SVM) |

|               |   |                                  |  |
|---------------|---|----------------------------------|--|
| Hybrid Method | Consistency,<br>Efficient,<br>Flexible,<br>Adaptability | Higher accuracy,<br>more Complex |  |
|---------------|---|----------------------------------|--|

### 3.4 Different techniques based on DL for HSD

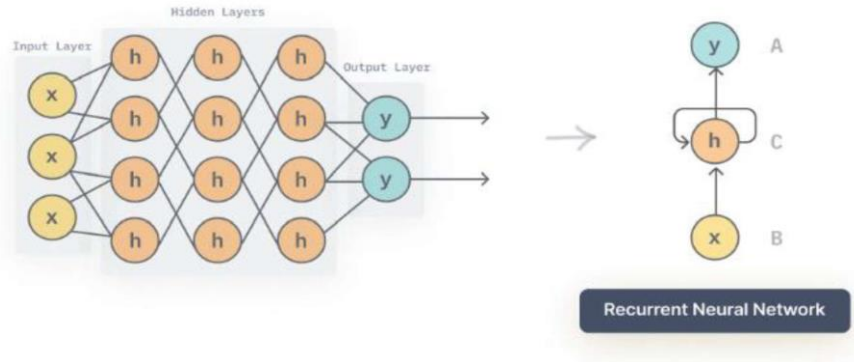
Various DL techniques have been utilized for hate speech detection, each providing unique advantages and methodologies for identifying and addressing harmful content online. Some prominent DL methods for HSD include:

**3.4.1 CNNs:** Primarily evolved for image processing tasks, CNNs have played a key role in text classification, including HSD. By applying filters over input text, CNNs seeks local patterns and features, enabling effective identification of hate speech based on linguistic cues.



**Figure 3.1 CNN**

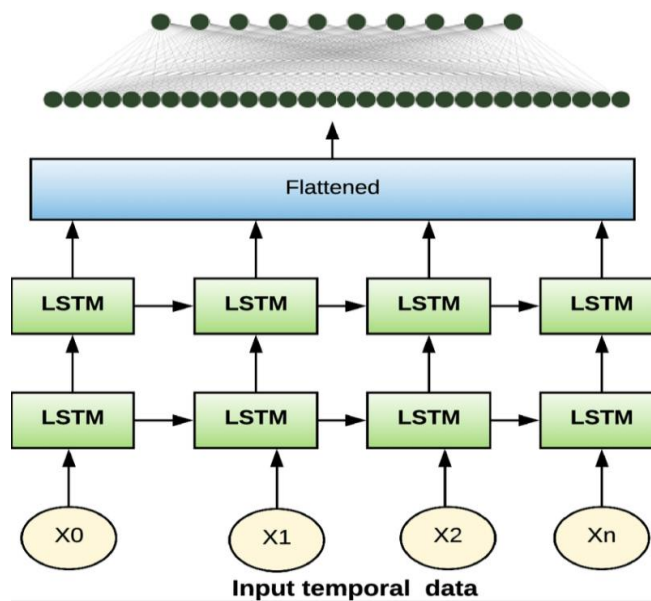
**3.4.2 RNNs:** Designed to manipulate sequential data, RNNs retain the memory of previous inputs, making them suitable for tasks involving text data like hate speech detection. They excel in capturing the contextual nuances of language, pivotal for the convention the meaning of the text.



**Figure 3.2 RNN**

**3.4.3 Long Short-Term Memory Networks (LSTMs):** LSTMs, the subtype of RNNs, compete in learning long-range dependencies in sequential data, overcoming issues like vanishing gradients. This makes them adept at seeking contextual information in text data, enhancing HSD capabilities.

LSTM [30], developed by Hochreiter and Schmidhuber, has found several applications, particularly in HS recognition, where it has been primarily adopted by IBM. One of the popular DL approaches for HSD. This architecture assimilates a memory component known as a cell, capable of retaining its value over time and maintaining it as input for subsequent operations.



**Figure 3.3 Architecture of LSTM**

Three gates within the cell keep an eye on the flow of data into and out of the structure:

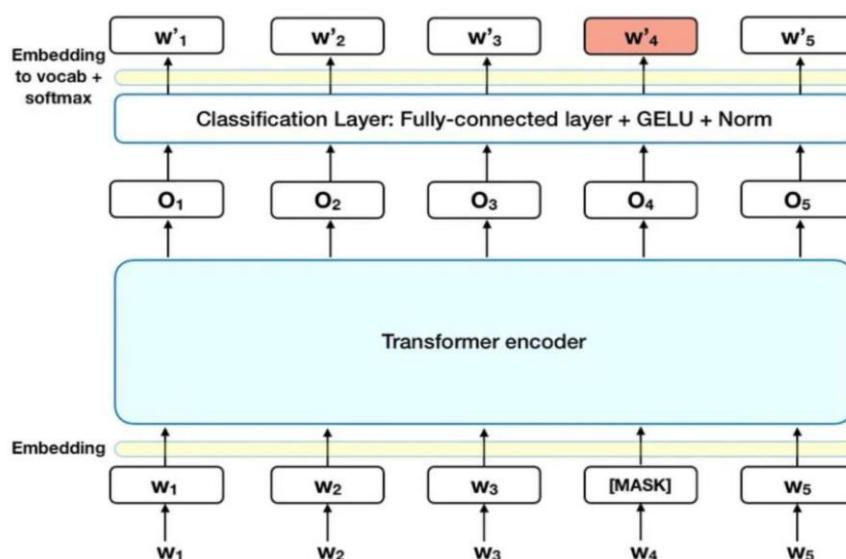
- The I/P gate controls the influx of new data into the memory unit.
- The forget gate, as its name suggests, manages the removal of obsolete information from the cell, expediting the storage of new data.
- The O/P gate oversees information within the given cell, determining the output of the cell.

**3.4.4 Transformers:** Transformers, renowned for their attention mechanisms, capture global dependencies in text data, making them popular in NLP tasks. Models like BERT, GPT derived from transformers, have been profitably deployed to HSD, achieving state-of-the-art results.

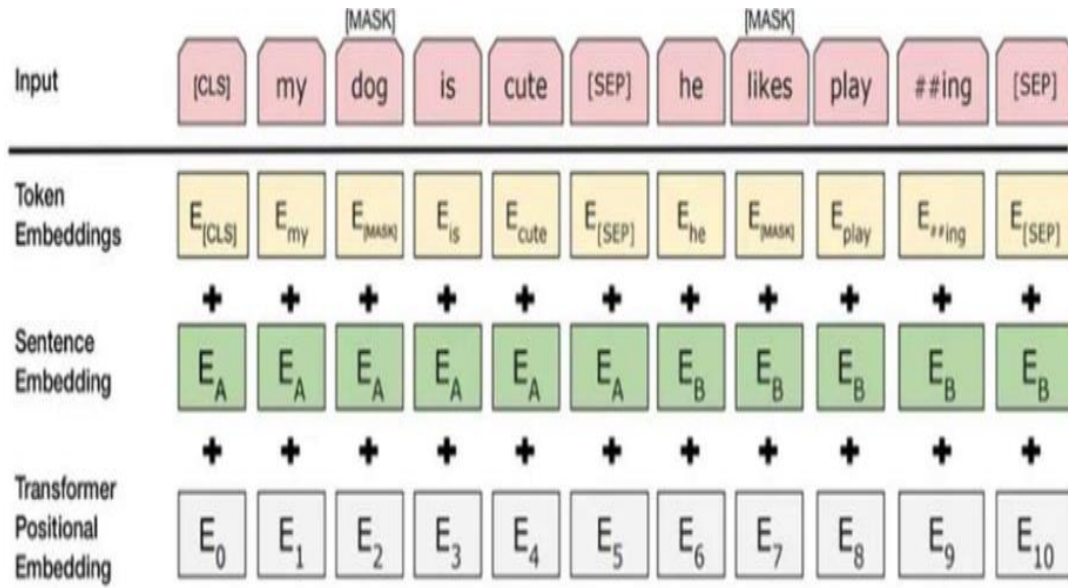
BERT is popular pretrained NLP models consist of special feature known as bi-directional, i.e. it can scan text from L-R and R-L. Since, BERT captures text from both the directions hence it is the most efficient technique to understand the meaning of the sentence. It uses transformer- based architecture for introducing parallelism which makes processing of the data efficient.

BERT is basically pre-trained using the following:

- MLM (Masked Language Model)
- NSP (Next Sequence Prediction)



**Figure 3.4.1** MLM (Masked Language Model)

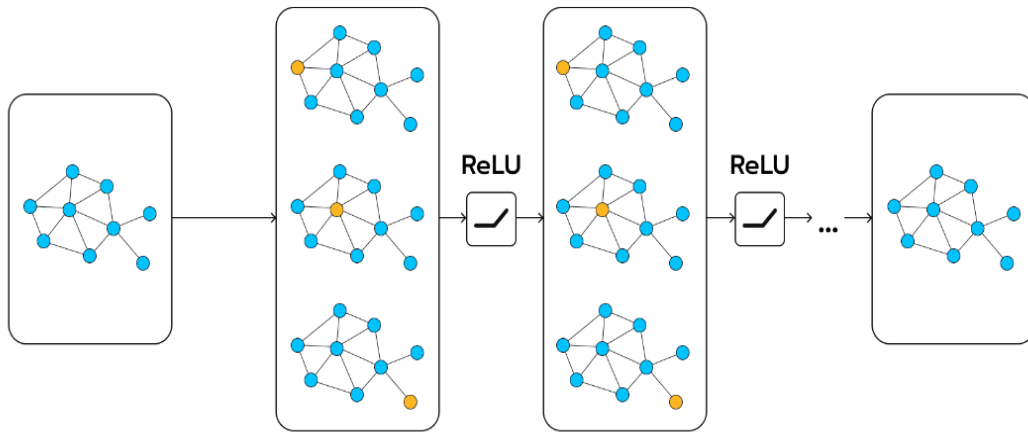


**Figure 3.4.2 NSP (Next Sequence Prediction)**

**3.4.5 Graph Convolutional Networks (GCNs):** Extending CNNs to graph-structured data, GCNs are suitable for tasks involving relational data like social networks. In hate speech detection, GCNs utilize the network structure of online communities to analyse patterns of hate speech propagation.

Graph Convolutional Networks (GCNs) are a highly potent multi-layer network architecture designed for machine learning tasks on graph-structured data [29]. GCN follows the following properties [28]:

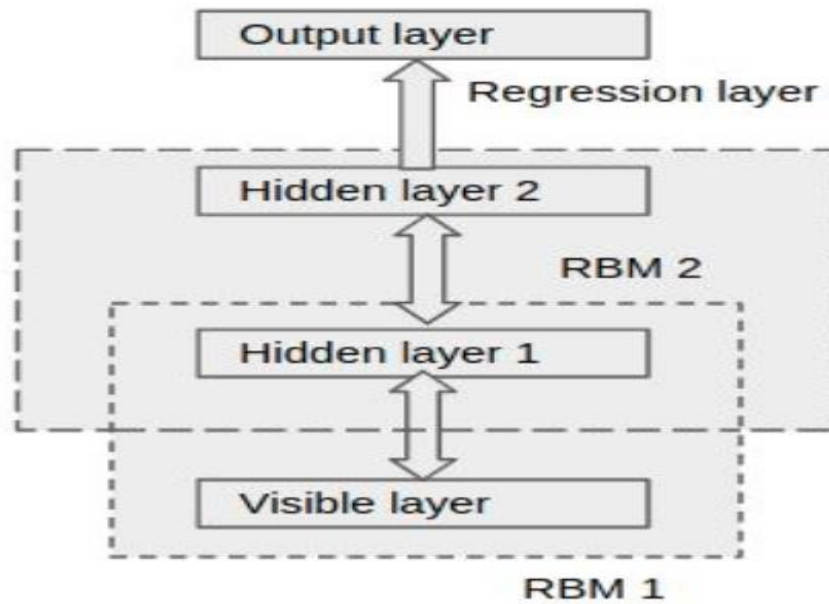
- (a) Nodes(n): The graph comprises nodes representing words and tweets, with the total number of nodes equal to the sum of tweets and the total count of unique words in the dataset.
- (b) Edges(e): The graph is defined by two distinct types of edges: tweet-word edges, which are constructed based on word occurrences within sentences, and word-word edges, which emerge from word co-occurrences within the dataset.
- (c) Weights(w): These can be determined using word embedding techniques such as TF-IDF considering the usefulness of the word and other relevant properties in the context of the dataset.



**Figure 3.5** GCN architecture

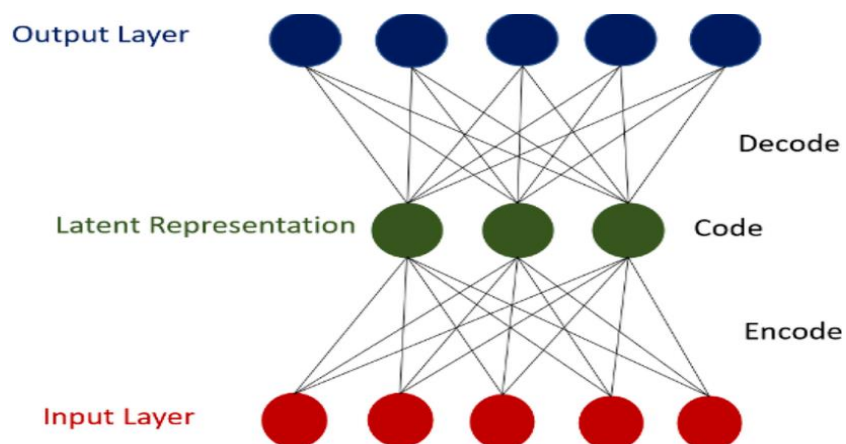
**3.4.6 Ensemble Learning:** It consolidates numerous base models to enhance overall prediction accuracy. By leveraging methods like bagging, stacking, and boosting ensemble learning improves the execution of deep learning models for hate speech detection, exploiting the strengths of diverse architectures.

**3.4.7 RBM:** RBM stands for Restricted Boltzmann Machine, a form of NN comprising 2 layers: visible layer (VL) & hidden layer (HL). Unlike traditional neural networks, RBMs lack connections within each layer but feature connections are established between the VL and HL. The objective of RBMs is to escalate the predicted data log probability. The inputs to RBMs are typically binary vectors, as they are learned from Bernoulli distributions. Activation functions in RBMs are calculated similarly to those in regular neural networks, typically using the logistic function to produce values between 0 and 1. Neurons in RBMs activate if their activation exceeds a random variable threshold. Visible units serve as inputs to hidden layer neurons, while hidden layer neurons compute probabilities based on the inputs received from the visible layer [26].



**Figure 3.6** RBM architecture

**3.4.8 AE:** Autoencoder (AE) is a conventional feedforward neural network crafted to extract a compact and widely distributed representation of a dataset. Comprising three layers, it is trained to reestablish inputs as outputs, enabling it to grasp certain features that efficiently capture variability within the dataset. When linear activation functions are applied, an AE can effectively perform dimensional reduction, akin to the functionality of Principal Component Analysis (PCA). After training, the activations of the hidden layer serve as learned features, rendering the top layer redundant [26].

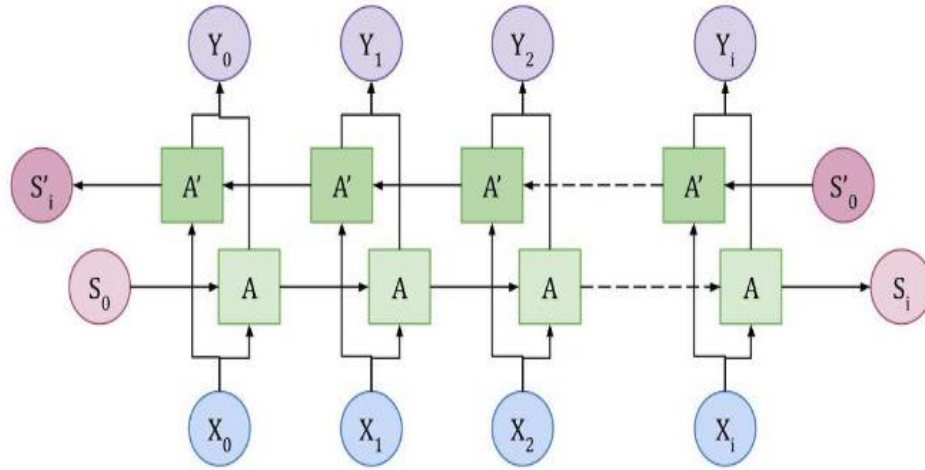


**Figure 3.7** Simple AE architecture



**3.4.9 Bi-LSTM:** It stands out as a variant of LSTM architecture extensively applied in NLP, particularly in tasks like HSD, Sentiment analysis, etc. Unlike conventional LSTMs, which analyse input sequences bi-directionally. This dual direction processing enables the model to assimilate contextual cues from both preceding and succeeding inputs, enhancing its ability to comprehend and interpret the sequential characteristics of text data.

In the context of my research, this method demonstrates the efficiency in capturing extended dependencies and subtle contextual nuances within textual content. This attribute makes them particularly adapt at discerning nuanced linguistic patterns indicative of hate speech. Leveraging information from both preceding and succeeding words in a sentence, Bi-LSTMs effectively grasp the syntactic and semantic structure inherent in hate speech corpus.



**Figure 3.8** Architecture of Bi-LSTM

## **CHAPTER 4**

### **LITERATURE SURVEY**

We studied different papers based on this topic. Identifying hate speech is difficult as some words have several meanings. Coded words and abbreviations are difficult to detect. We have to depend on long input data to understand the actual context.

Skip gram models and continuous bag of words plays a crucial role in this segment. Detecting context from both direction is more efficient.

#### **4.1 Literature Survey of Deep Learning Models**

This paper offers an understanding of the obstacles encountered in hate speech detection and introduces a dataset tailored for training and assessing hate speech detection models. It establishes the foundation for further investigations into automated hate speech detection [31]. This study presents an approach to detecting HS in OSN using ML techniques. It discusses the importance of feature engineering and model selection in improving detection accuracy [32]. This paper examines the predictive elements utilized for finding HS on Twitter, with a certain prominence on lexical, syntactic, and semantic indicators. It underscores the significance of contextual details and adapting to different domains in the process of hate speech detection [12]. The workshop paper rousts into the difficulties associated with distinguishing between profanity and hate speech in online content. It explores the constraints of current hate speech detection approaches and suggests potential directions for future research [33].

This pervasive survey offers a broad examination of automatic HSD techniques, enveloping rule-based systems, ML methodologies, and DL approaches. It gauges the merits and obstacle of each method while also pinpointing areas of ongoing research and unresolved challenges within the field [34]. This survey figure out recent progress in HSD through the lens of NLP techniques. It envelops a spectrum of feature representations, classification algorithms, and evaluation metrics employed in hate speech detection tasks [35]. Additionally, this systematic survey delivers a thorough outline of hate speech detection methodologies, categorizing them according to their underlying approaches and feature representations. It assesses the efficacy of various methods and highlights avenues for future research [36]. Furthermore, this review article consolidates

the latest advancements in HSD, focusing on DL techniques. It rousts into the architectures, training methodologies, and performance metrics utilized in deep learning-driven hate speech detection systems [37]. This concentrated on terms sexism & racism: aiming to identify hate speech occurrences on X, SVM, a supervised learning algorithm give good result with F1-score up to 80%.

This paper proposed a muti-class classifier to divide the language into the hate, offensive and non-offensive languages. This paper to classify the tweets a naïve base approach is used. LR gives outstanding results for detection of detest words and get precision up to **90%**. Word2vec word embeddings and CNN gives fantastic results. Bi-LSTM along with word2Vec gets an accuracy of 91.10%. Got great results in detection of hate speech in multilingual using MBERT. Created a deep NN for hate speech with target of 92% F1-score. This paper utilized RNN to incorporate unigram, bigram for character embeddings. The proposed method entailed building a model for text sentiment analysis that combined a bi- directional gated recurrent unit (GRU) and MCCNN). Using IMDB dataset, achieved an accuracrate of 91.20% with success.

CNN-GRU model performed well on both public and private dataset. This paper uses RNN to detect difference between numerous hatred words. It used n-gram (1 gram) model with SVM for classification between non-hate & hate for first data corpus and second dataset for classification, identifying content between clean, offensive & hate with accuracy of 87.4% and 78.4% respectively. The fusion-oriented model, which merged 3 CNN, yielded an average of 75.4%-Score of accuracy.

## **4.2 Literature Survey of various GCN Models**

Initially, integrating Graph Neural Networks into abusive language detection involved combining Graph Convolutional Networks (GCN) with Logistic Regression (LR), resulting in superior performance compared to other LR or LR hybrid models [38]. A novel Graph Convolutional Network (GCN) classifier, SOSNet, was introduced, utilizing a graph generated from thresholded cosine similarities among tweet embeddings. By leveraging a dataset enhanced with Dynamic Query Expansion (DQE), the study evaluated the effectiveness of the proposed GCN model against eight tweet embedding techniques and six alternative classification models across datasets of varying sizes. Results demonstrated that the GCN model performed comparably or better than the

baseline models, with the combination of SOSNet and SBERT achieving the highest accuracy and F1-score [39].

In the coarse-grained assessment, the Relational Graph Convolutional Network (RGCN) demonstrated superiority in identifying bogus posts with an F1 score of 0.97. The most encouraging outcomes were obtained by combining RGCN and BERT, highlighting the need of using semantic meaning in addition to contextual information to improve classification accuracy [40] in multiple datasets across multilingual domain.

GCN's viability as an estimator and data-efficient solution was investigated using the bi-Courage model. By combining two text-to-graph algorithms with different modelling approaches, biCourage achieves better performance than basic BERT in every way [41]. Using Twitter data, a dual-layer GCN was created to improve the classification accuracy of several models. After conducting a thorough comparative analysis using nine supervised classical machine learning algorithms for classification, the GCN model outperformed all other traditional classifiers.

HateNet, a Graph Convolutional Network (GCN) model, combined with a hybrid (semi-supervised) approach employing subDQE was introduced across three distinct datasets. The subDQE augmentation improved classification results, with the HateNet + SBERT configuration outperforming all machine learning techniques and models [43]. SyLSTM, a recently proposed model, combined syntactic features from the dependency parse tree of a sentence and semantic attributes from word embeddings within deep learning frameworks utilizing Graph Convolutional Networks (GCN). This approach significantly outperformed the state-of-the-art BERT model while requiring considerably fewer parameters, leading to less training time compared to BERT [44].

DGCSKT merged DGC and Sentiment Knowledge Transfer, utilizing SDG and DGC operations to reinforce contextual information comprehension. It outperformed BERT, Bi-LSTM, and SKS in both accuracy and F1 score [45].

The BERT-CNN model exhibited admirable performance compared to the simple 2-layer GCN model. Further enhancements in results were observed by integrating additional hidden layers and utilizing a more complex graph structure [46]. HA-GCEN, a model

constructed with hypergraph convolution layers, demonstrated significant performance improvements and benefits in terms of model development and training complexity. It also exhibited a notable increase in both recall and precision, consequently leading to an enhancement in F1-score [47]. This paper conducts a sweeping comparison of various techniques entrenched performance evaluation. The analysis aims to provide a detailed examination of different models and their effectiveness in tackling the problem at hand [48].

## **CHAPTER 5**

### **MODEL ANALYSIS**

In this segment of the thesis, I analyzed the different existing models based on GCN for HSD and tried to present a summary of each model based on various parameters.

- The **LR+ GCN** model not only captures the structural patterns of online social discourse but also the linguistic behaviour of users. Essentially based on LR+Auth, it integrates GCN for classification. Comparative analysis with LR, LR+Auth, and a basic GCN model demonstrates its superior performance in overall analysis. However, there is a decline in performance for specific tags such as racism.
- The **SOSNet** model is focused on three core notions. First, it includes the formulation of an online Dynamite Query Expansion operation by utilizing merged keyword browsing. Secondly, it erects a diagram frame of tweet embarks and utilizes a Graph Convolutional Net for accurate cyberbullying ranking. Thirdly, it gives importance to the establishment of a balanced multiple-class cyberbullying dataset from DQE and its national disclosure. When matched with current ranking formulas and screwing techniques, applying SOSNet with SBERT stands away as the most promoting tactic, reaching an amazing F1 value of more to 92%.
- Another model, **biCourage**, explores GCN models by extending the MeanPool model introduced by Wilkens and Ognibene [49]. It incorporates normalization in each layer of GCN and substitutes GCN layers with GraphSAGE layers, which closely resemble the GCN layer [50]. GraphSAGE aggregates information from local neighbours, and as the process iterates, nodes incrementally acquire information from other parts of the graph [50] and this model surpasses Bert's fine-tuning process and offers a training speed that is 3.85 times faster.
- **HateNet**, an advanced model, expands upon the principles of SOSnet by integrating a semi-supervised hybrid approach known as subDQE, which combines substitution-based augmentation with DQE. This model consists of four primary components: a semantic similarity graph, weighted DropEdge, short text data augmentation, and graph convolution.
- The **SyLSTM** model is composed of six key components: input tokens, BiLSTM layer, GCN layer, embedding layer, feed-forward layer, and output layer. Interestingly,

when leveraging pretrained Glove embeddings, the SyLSTM model demonstrates noteworthy performance, even exceeding that of the basic SyLSTM model.

- The most updated **Hyperedge-Abundance Graph Convolutionally Enhanced Networking** model utilizes hypergraph convolutional layers to capture underlying data patterns. Through investigating non-regional data and high-order correlations within the hypergraph, HA-GCEN utilizes the "enhanced connection" to perform non-linearity mapping on those features. The model's capability to handle high-dimension data comes from its integrated approach, combining hypergraph convolution layers and the "enhanced connection," thereby enhancing discriminative ability and generalization performance.

## CHAPTER 6

### RESULT AND DISCUSSION

The subsequent section provides assess how different models perform using various metrics, all mixed up with big words and stuff.

#### **6.1 Performance metrics**

Performance metrics are key tools for evaluating the effectiveness and accuracy of different models, even the ones used in love speech detection. These metrics provide some numericals to show good or bad a model is by comparing its guesses with some labels. Some common performance metrics used in love speech detection include:

a. **Accuracy:** it shows how precise a model is at guessing, which is usually shown as the part of correct guesses compared to all guesses.

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{N}} \dots\dots\dots (\text{i})$$

b. **Precision:** "Precision" measures the model's ability to accurately recognize positive instances among all guesses of positiveness, which is some math thing showing how well the model picks out the right stuff among its positive guesses.

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \dots\dots\dots (\text{ii})$$

c. **Recall:** it checks how well the model can figure out all the actual positive stuff among all the positive stuff in the data.

$$\text{Rec} = \frac{\text{TP}}{\text{TP} + \text{FN}} \dots\dots\dots (\text{iii})$$

d. **F1 Score:** The F1 score gives a fair look at a model's performance by mixing precision and recall. It figures out some stuff of average of precision and recall, treating both as equals.

$$\text{F1} = \frac{2 * \text{Rec} * \text{Pre}}{\text{Rec} + \text{Pre}} \dots\dots\dots (\text{iv})$$

#### **6.2 Comparison of various deep learning technique(s)**

The subsequent section performs the comparative analysis of different deep learning technique(s) utilizing a variety of performance metrics.

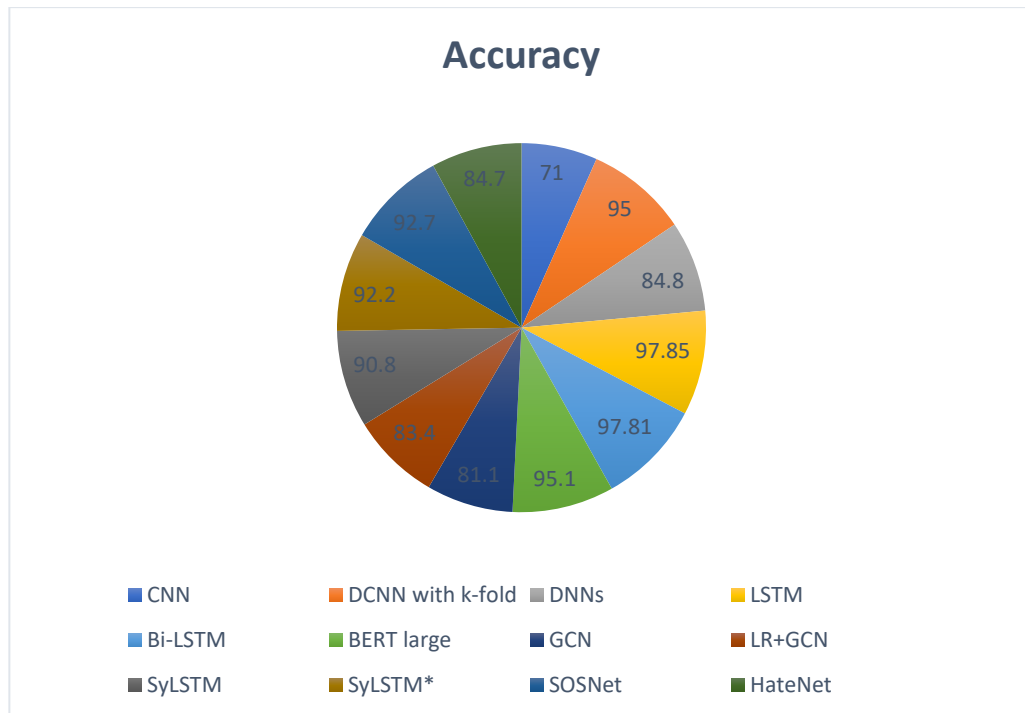
The table extant the numerical values of various model, facilitating an analysis of the performance.



**Table 6.1:** Showing various parameters in percentage

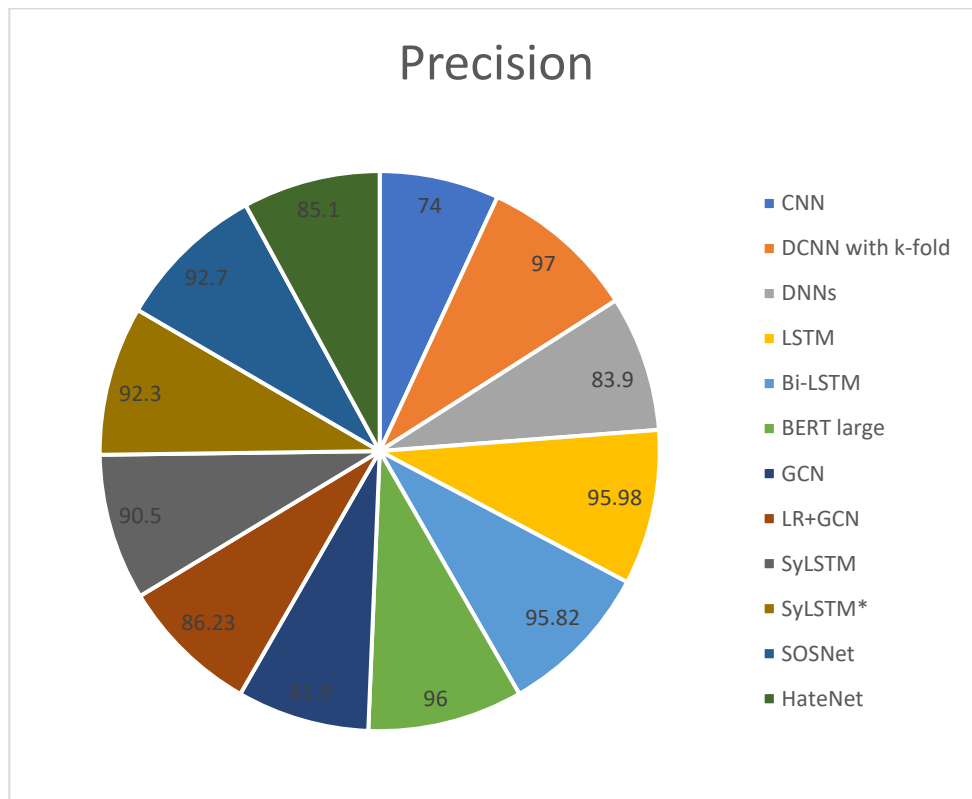
| Models                | Accuracy | Precision | Recall | F1 Score |
|-----------------------|----------|-----------|--------|----------|
| CNN [52]              | 71.0     | 74.0      | 82.0   | 86.0     |
| DCNN with k-fold [53] | 95.0     | 97.0      | 88.0   | 92.0     |
| DNNs [48]             | 84.8     | 83.9      | 84.0   | 83.9     |
| LSTM [51]             | 97.85    | 95.98     | 99.86  | 97.85    |
| Bi-LSTM [51]          | 97.81    | 95.82     | 99.90  | 97.81    |
| BERTlarge [54]        | 95.1     | 96.0      | 96.0   | 96.46    |
| GCN [2]               | 81.1     | 81.90     | 79.42  | 80.56    |
| LR+GCN [2]            | 83.4     | 86.23     | 84.73  | 85.42    |
| SyLSTM [44]           | 90.8     | 90.5      | 91.4   | 91.4     |
| SyLSTM* [44]          | 92.2     | 92.3      | 92.8   | 92.7     |
| SOSNet [32]           | 92.7     | 92.7      | 92.4   | 92.58    |
| HateNet [43]          | 84.7     | 85.1      | 84.2   | 84.3     |

Figure 6.1 exhibit pie chart plotted to spectacle accuracy of various models for HSD. We can see that on the given dataset, CNN shows the minimum accuracy while the LSTM is superior and achieved accuracy of 97.85%.



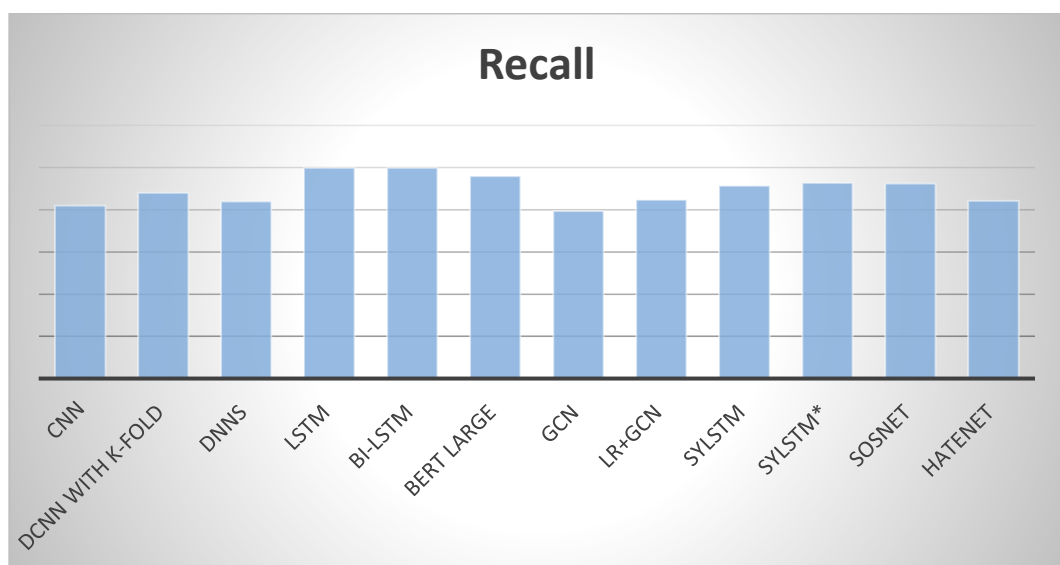
**Figure 6.1** Accuracy chart of DL techniques

Figure 6.2 displays pie chart plotted to spectacle the precision of various models for HSD. We can see that on the given dataset CNN shows the minimum precision while the DCNN with K-fold possess maximum precision of 97%.



**Figure 6.2** Precision Chart of various DL techniques

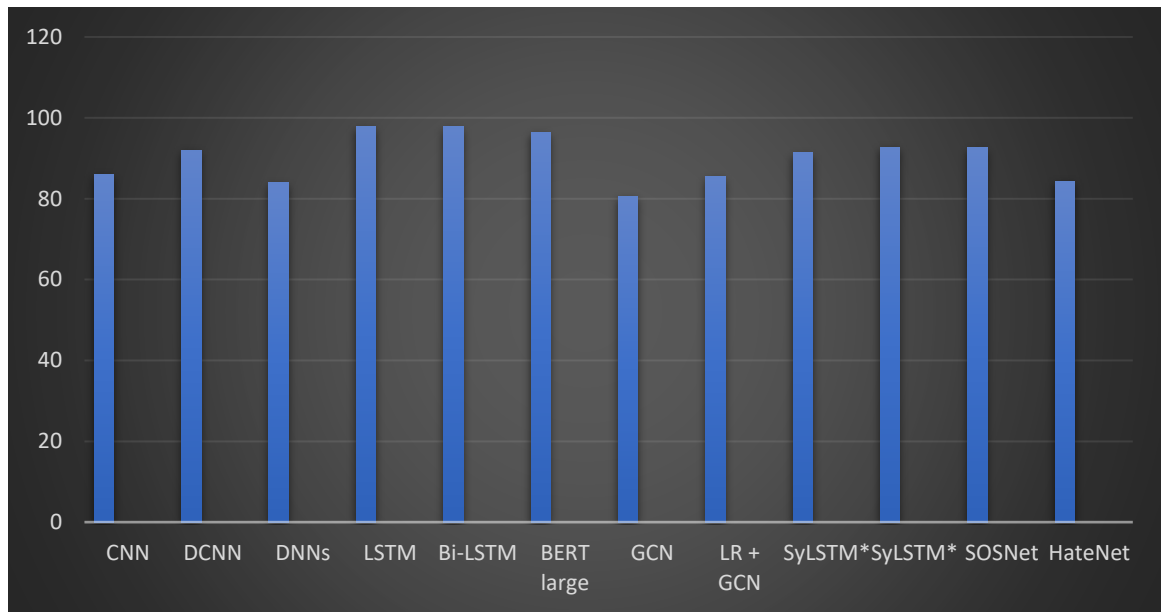
Figure 6.3 displays the recall of various models. Bi-LSTM achieves the highest recall value while GCN have lowest recall values.



**Figure 6.3** Recall chart of DL techniques

Additional visual analysis will be carried out through the utilization of diverse charts. For instance, Figure 6.4, which depicts a line graph showcasing the F1 scores for LSTM, CNN, DNN, DCNN, Bi-LSTM, GCN, LR+GCN, SyLSTM, SOSNet, and HateNet, demonstrates that deep learning methodologies achieve superior F1 scores, with six models exceeding the 90% mark.

In contrast, GCN and DNN exhibit the lowest F1 scores, whereas LSTM and Bi-LSTM demonstrate the highest values, underscoring the variability in performance across different deep learning architectures



**Figure 6.4** Graph showing F1 Score of various models

### 6.3 Comparative Analysis of various GCN models

For each model, we will assess various parameters and compare their performance. Table 3 will present the performance metrics of various models. The HA-GCEN model is particularly noteworthy, showing a significant increase of 10-15% in both F1 score and precision across different datasets. It also surpasses other models like GCN, RSGNN, HateBert, among others. Combining Sentence BERT with HateNet leads to outperformance in provision of accuracy and precision as to many related models.

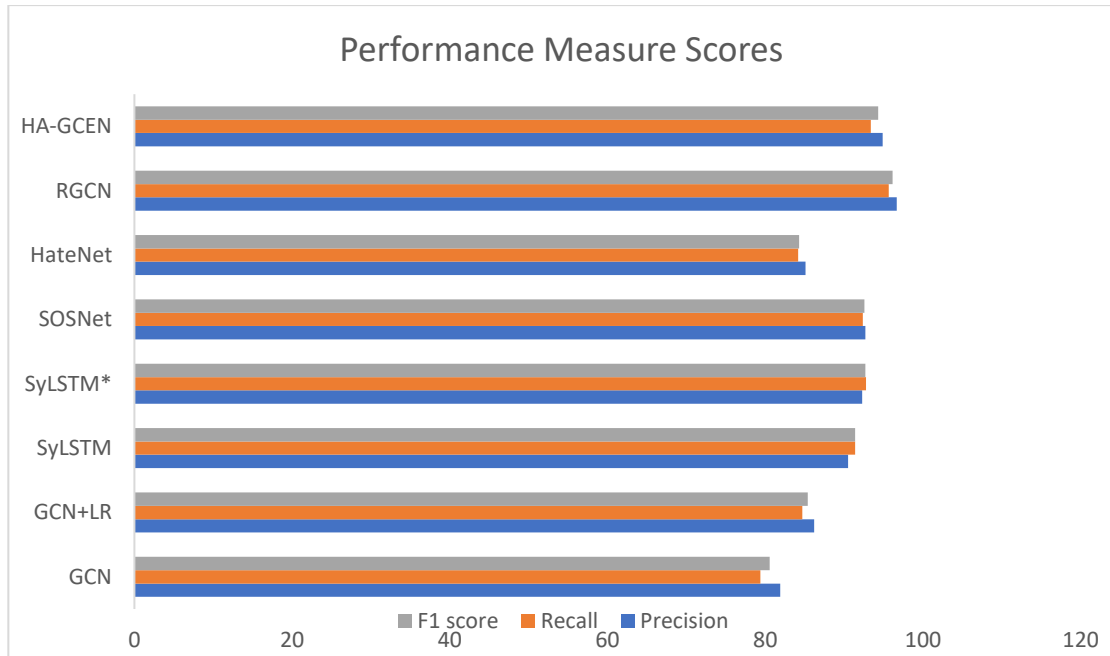
The GCN model achieves an F1 score of up to 80.56% with precision of 81.90 while LR+GCN further boosts it to 85.42%. DGCSKT demonstrates robust performance across various datasets, showcasing an increase in accuracy and F1 score by 3.38% and 3.88%, respectively, compared to existing models. Moreover, leveraging SyLSTM with Glove

embeddings elevates its F1 score to 92.7%. When integrated with SBERT, SOSNet emerges as a robust contender, achieving an accuracy of up to 92.7% and an F1 score of up to 92.58%.

Moreover, RGCN + BERT outperforms other models when the data are segmented, achieving an impressive F1 score of up to 96.614%. biCourage is yet another model works well on various languages subtasks.

**Table 6.2:** Performance evaluation

| Model   | Precision | Recall | F1 score |
|---------|-----------|--------|----------|
| GCN     | 81.90     | 79.40  | 80.56    |
| GCN+LR  | 86.23     | 84.73  | 85.42    |
| SyLSTM  | 90.5      | 91.4   | 91.4     |
| SyLSTM* | 92.3      | 92.8   | 92.7     |
| SOSNet  | 92.7      | 92.4   | 92.58    |
| HateNet | 85.1      | 84.2   | 84.3     |
| RGCN    | 96.7      | 95.68  | 96.14    |
| HA-GCEN | 94.9      | 93.39  | 94.34    |



**Figure 6.5** Performance scores of various GCN models

## **CHAPTER 7**

### **CONCLUSION**

The epidemic rise in the use of HS on the OSN creates a critical challenge for society. Multiple groups belonging to particular sections of society have been targeted based on discrimination. In recent times, DL models have emerged as highly effective tools for HSD, which outplay traditional approaches used earlier. Among these, LR shows good results, but LSTM and bi-LSTM modeling have demonstrated exceptional performances, even though facing challenges such as handling multilingual datasets and classification issues. BERT-based modeling has also exhibited impressive results, excelling in English and detecting HS in languages like Arabic, Bengali, and Marathi. Additionally, the introduction of GNN based GCN models has shown promising results in HSD.

This comparative analysis of various DL technologies provides valuable insights for enhancing haters and HSD and serves as a foundation for future research in these fields. Our analysis results will assist as a valuable resource for researchers, practitioners, and policymakers involved in combating online HS. By understanding the virtue of multiple DL technologies, stakeholders can make informed decisions about the selection and implementation of most appropriate models for HSD. These steps will solely depend on the model analysis. Moving forward, further R&D efforts are certified to address the prevailing challenges and further enhance the efficiency of HSD systems in cherishing a safer online environment.

While various methods have been deployed for HSD in recent years, model using GCN has exhibited significant performance compared to existing ones. While GCN models generally outplay other classifiers when compared in various datasets, they still face several limitations. Challenges such as multilingual detections, data categorization issues, and capturing long-distance semantics remain significant hurdles in these domains. Moreover, it's observed that GCN models are computationally constrained (more costly) when compared to BERT, primarily due to the larger memory requirements for storing graphs used in GCN models. Given these limitations, further research is required to minimize the storage and enhance the capabilities of GCN modeling for broader applications in HSD.

## **REFERENCES**

- [1] Wright M. F, Harper B. D, and Wachs S, (2019) “The associations between cyberbullying and callous-unemotional traits among adolescents: The moderating effect of online disinhibition,” *J. Personality Individual Differences*, vol. 140, pp. 41-45.
- [2] Davidson T, Warmesley D, Macy M, and Weber I, (2017) “Automated hate speech detection and the problem of offensive language,” in *Proc. ICWSM*, 2017, pp. 1–4.
- [3] Gershgorn D and Murphy M, (2017). Facebook is Hiring More People to Moderate Content than Twitter has at Its Entire Company Quartz. Accessed: Jun. 20, 2019. [Online]. Available: <https://bit.ly/2ZbhsHu>
- [4] Vega T, (2019) “Facebook says it failed to bar posts with hate speech,” *The New York Times*, 2013. Accessed: Jun. 10, 2019. [Online]. Available: <https://nyti.ms/2VXy9Ex> .
- [5] Meyer R, (2019) “Twitter’s famous racist problem,” *The Atlantic*, 2016. Accessed: Jul. 5, 2019. [Online]. Available: <https://bit.ly/38EnFPw>.
- [6] Rodríguez-Sánchez F, Carrillo-de-Albornoz J and Plaza L, (2020) "Automatic Classification of Sexism in Social Networks: An Empirical Study on Twitter Data," in *IEEE Access*, vol. 8, pp. 219563-219576. doi: 10.1109/ACCESS.2020.3042604.
- [7] Shi Z. Ryan, Wang C, and Fang F, (2020) “Artificial intelligence for social good: A survey” *arXiv:2001.01818*. [Online]. Available: <http://arxiv.org/abs/2001.01818>
- [8] Khatua A, Cambria E, and Khatua A, (2018) “Sounds of silence breakers: Exploring sexual violence on Twitter,” in *Proc. ASONAM*, pp. 397–400.
- [9] Khatua A, Cambria E, Ghosh E, Chaki N, and Khatua A, (2019) “Tweeting in support of LGBT? A deep learning approach,” in *Proc. ACM India Joint Int. Conf. Data Sci. Manage. Data*, pp. 342–345.
- [10] Cambria E, Chandra P, and Hussain A, (2010) “Do not feel the trolls,” in *Proc. SDoW Workshop 9th Int. Semantic Web Conf*, pp. 1–12.
- [11] Ji S, Pan S, Li X, Cambria E, Long G, and Huang Z, (2020) “Suicidal ideation detection: A review of machine learning methods and applications,” *IEEE Trans. Comput. Social Syst.*, early access. doi: 10.1109/TCSS.2020.3021467.
- [12] Waseem Z, (2016) “Are you a racist or am i seeing things? Annotator influence on hate speech detection on Twitter,” in *Proc. 1st Workshop NLP Comput. Social Sci* pp. 138–142.
- [13] Waseem Z and Hovy D, (2016) “Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter,” in *Proc. NAACL Student Res. Workshop*, pp. 88–93.
- [14] Xiang G, Fan B, Wang L, Hong J, and Rose C, (2012) “Detecting offensive tweets via topical feature discovery over a large scale Twitter corpus,” in *Proc. 21st ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, pp. 1980–1984.
- [15] Gianfortoni P, Adamson D, and Rose C, (2011) “Modeling of stylistic variation in social media with stretchy patterns,” in *Proc. EMNLP*, pp. 49–59.
- [16] Cambria E, (2016) “Affective computing and sentiment analysis,” *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 102–107.

- [17] Vigna F. Del, Cimino A, Dell'Orletta F, Petrocchi M, and Tesconi M, (2017) "Hate me, hate me not: Hate speech detection on facebook," in Proc. ITASEC, 2017, pp. 86–95.
- [18] Gitari N.D, Zhang Z, Damien H, and Long J, (2015) "A lexicon-based approach for hate speech detection," Int. J. Multimedia Ubiquitous Eng., vol. 10, no. 4, pp. 215–230.
- [19] Tang Y and Dalzell N, (2015) "Classifying hate speech using a two-layer model," Statist. Public Policy, vol 6, no. 1, pp. 80-88. Doi 10.1080/2330443X.2019.1660285.
- [20] Badjatiya P, Gupta S, Gupta M, and Varma V, (2017) "Deep learning for hate speech detection in tweets" on Proc. ACMWWW, pp. 759-760.
- [21]. Waseem Z, (2016)" 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, pages 138– 142.
- [22]. Waseem Z, and Hovy D (2016) "Hateful symbols or hateful people? predictive features for hate speech detection on twitter" in Proceedings of the NAACL student research workshop,88-93.
- [23]. Davidson T, Warmesley D, Macy M, and Weber I (2017) "Automated hate speech detection and the problem of offensive language" in Eleventh international aaai conference on web and social media.
- [24]. Vu X.S, Vu T, M.-V. Tran, Le-Cong T, and Nguyen H.T.M (2019) , "HSD shared task in VLSP campaign 2019: Hate speech detection for social good," in Proceedings of VLSP 2019.
- [25] Femi Emmanuel Ayo, Olusegun Folorunso, Friday Thomas Ibharalu, Idowu Ademola Osinuga, Machine learning techniques for hate speech classification of twitter data: State-of-the-art, future challenges and research directions, Computer Science Review, Volume 38, 2020, 100311, ISSN 1574-0137, <https://doi.org/10.1016/j.cosrev.2020.100311>.
- [26] V. Pream Sudha and R. Kowsalya, (2015) "A Survey on Deep Learning Techniques, Applications and Challenges", International Journal of Advance Research in Science and Engineering (IJARSE), Vol. No.4, Issue 03, March 2015, pp. 311-317.
- [27] Shaveta Dargan, Munish Kumar, Maruthi Rohit Ayyagari and Gulshan Kumar, (2019) "A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning", Archives of Computational Methods in Engineering, pp. 1-23. <https://doi.org/10.1007/s11831-019-09344-w>.
- [28] SS. Pandey, (2023) "A Comparative study of BERT-CNN and GCN for Hate Speech Detection".
- [29]. Kipf T.N, and Welling M (2016) "Semi- supervised classification with graph convolutional networks" arXiv preprint [online]. Available: <http://arxiv.org/abs/1609.02907>.
- [30] D. Goularas and S. Kamis, "Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data," 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), Istanbul, Turkey, 2019, pp. 12-17, doi: 10.1109/Deep-ML.2019.00011.
- [31]. Davidson, T., Warmesley, D., Macy, M., & Weber, I. (2017). Automated Hate Speech Detection and the Problem of Offensive Language. In Proceedings of the 11th International AAAI Conference on Web and Social Media (pp. 512-515).
- [32]. Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., & Chang, Y. (2016). Abusive Language

Detection in Online User Content. In Proceedings of the 25th International Conference on World Wide Web (pp. 145-153).

[33]. Malmasi, S., & Zampieri, M. (2018). Challenges in Discriminating Profanity from Hate Speech. In Proceedings of the Second Workshop on Abusive Language Online (pp. 3-11).

[34] Fortuna, P., Nunes, S., & Bhat, S. (2018). A Survey on Automatic Detection of Hate Speech in Text. *ACM Computing Surveys*, 51(4), 1-30.

[35]. Zhang, Y., & Luo, T. (2020). A Survey on Hate Speech Detection using Natural Language Processing. *Information Processing & Management*, 57(2), 102025.

[36]. Mishra, R., & Bhattacharyya, P. (2020). A Systematic Survey of Hate Speech Detection Techniques. *ACM Computing Surveys*, 53(4), 1-39.

[37]. Alambo, A., & Imran, M. (2021). Deep Learning-Based Approaches for Hate Speech Detection: A Review. *Journal of Information Science*, 47(3), 315-335.

[38]. Mishra P, Tredici M, Yannakoudakis H, and Shutova E (2019) "Abusive Language Detection with Graph Convolutional Networks" in proceedings of NAACL-HLT 2019, pages 2145-2150.

[39]. Wang J, Fu K, and Lu C (2020) "SOSNet: A Graph Convolutional Network Approach to Fine-Grained Cyberbullying Detection", in IEEE international conference on Big Data. DOI: 10.1109/BigData50022.2020.9378065.

[40]. Sarthak, Shukla S, and ARYA K.V (2021)" Detecting Hostile Posts using Relational Graph Convolutional Network", in proceedings of arXiv:2101.03485v2.

[41]. Wilkens R, and Ognibene D (2021)" biCourage: ngram and syntax GCNs for hate speech detection" in forum for Information Retrieval Evaluation.

[42]. Saeidi M, Milios E, and Zeh N (2021)" Graph Convolutional Networks for Categorizing Online Harassment on Twitter" in 20<sup>th</sup> IEEE International Conference on Machine Learning and Applications (ICMLA).

[43]. Duong C, Zhang L, and Lu C (2022)" HateNet: A Graph Convolutional Network Approach to Hate Speech Detection" in proceedings of IEEE International Conference on Big Data.

[44]. Goel D, and Sharma R (2022) "Leveraging Dependency Grammar for Fine-Grained Offensive Language Detection using Graph Convolutional Networks", in arXiv:2205.13164v1.

[45]. Fan X, Liu J, Liu J, Tuerxun P, Deng W, and Li W, (2023) "Identifying Hate Speech Through Syntax Dependency Graph Convolution and Sentiment Knowledge Transfer", published in IEEE Access (volume 12) DOI: [10.1109/ACCESS.2023.3347591](https://doi.org/10.1109/ACCESS.2023.3347591).

[46]. Pandey S.S (2023) "A Comparative study of BERT-CNN and GCN for Hate Speech Detection".

[47]. Mu Y, Yang J, Li T, Li S, and Liang W (2023) "HA-GCEN: Hyperedge- Abundant Graph Convolutional Enhanced Network for Hate Speech Detection" available at SSRN: <https://ssrn.com/abstract=4677383> or <http://dx.doi.org/10.2139/ssrn.4677383>.

[48]. Yadav D, Sain M.K, and Abraham B (2023), "Comparative Analysis and Assessment of Different Hate speech Detection Learning Techniques" in *Journal of Algebraic Statistics* volume 14, No 1, p. 29-48.



- [49]. Wilkens R, Ognibene D, and Mb-courage@ exist (2021) “Gen classification for sexism identification in social networks” EXIST; sEXism Identification in Social neTworks- First Shared Task at IberLEF2021.
- [50]. Hamilton W.L, Ying, and Leskovec J (2017) “Inductive representation learning on large graphs” in Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 1025–1035.
- [51] Paul C, and Bora P (2021) “Detecting Hate Speech using Deep Learning Techniques,” Int. J. Adv. Comput. Sci. Appl., doi: 10.14569/IJACSA.2021.0120278.
- [52]. Rani P, Suryawanshi S, Goswami K, Chakravarthi B.R, Fransen T, and McCrae J.P (2020) “A Comparative Study of Different State-of-the-Art Hate Speech Detection Methods in {H}indi-{E}nglish Code-Mixed Data,” Proc. Second Work. Trolling, Aggress. Cyberbullying.
- [53]. Roy P.K, Tripathy A.K, Das T.K, and Gao X.Z (2020) “A framework for hate speech detection using deep convolutional neural network,” IEEE Access, doi: 10.1109/ACCESS.2020.3037073.
- [54]. Saleh H, Alhothali A, and Moria K (2023) “Detection of Hate Speech using BERT and Hate Speech Word Embedding with Deep Model” in proceedings of Applied Artificial Intelligence, 37:1, DOI: [10.1080/08839514.2023.2166719](https://doi.org/10.1080/08839514.2023.2166719).

## LIST OF PUBLICATIONS AND THEIR PROOFS

1. Anurag Sharma “Comparative Analysis of Deep Learning Techniques for Hate Speech Detection” Accepted at the “**2<sup>nd</sup> International Conference on Machine Intelligence for Research and Innovations-2024 (MAITRI)**” at National Institute of Technology, Srinagar.

Paper ID: 224





ANURAG SHARMA &lt;annuragsharma2311@gmail.com&gt;

## Acceptance Letter MAITRI-2024

1 message

Microsoft CMT &lt;email@msr-cmt.org&gt;

Sun, Apr 28, 2024 at 1:25 PM

Reply-To: Om P Verma &lt;vermaop@nitj.ac.in&gt;

To: Anurag Sharma &lt;annuragsharma2311@gmail.com&gt;

\*\*\*\*\*Read all the instructions carefully\*\*\*\*\*

\*\*\*\*\* Please ignore if already received \*\*\*\*\*

Dear Anurag Sharma,

We are glad to inform you that your manuscript with Title: Comparative Analysis of Deep Learning Techniques for Hate Speech Detection, having Paper ID: 224 has been accepted for the ORAL presentation (in hybrid Format) at MAITRI-2024. This means that your manuscript is among the top 25% of the manuscripts received/reviewed. Also, your manuscript is selected for publication in the conference proceedings published by LNNS Series of Springer subjected to the following conditions:

1. All the satisfactory modifications/revisions should be incorporated as suggested by the reviewer(s) comments.
2. The Editorial Board decides the final decision based on the revision or revised version of the paper received.
3. The best papers based on the review scores plus deep paper analysis will be awarded under different categories. The result of the same would be available on the conference website (Awards Sections). The award ceremony will be disclosed in due time. All winners of the best paper selections will get the certificate.
4. Use of Generative AI tools (like chatGPT, Grammarly, etc.) is less than 10% and has only been used to improve the clarity of the content.

Note: For the extended version(s) in Special Issue of Journals

5. The preference would be given to those registered authors who have submitted their paper in MAITRI-2023 with International collaborators (as a co-authors). However, the first author/Corresponding/Registered author may be an Indian.

\*\*\*\*Further, we would like to mention that we have very limited seats available, and once they are occupied, we shall not accommodate any further registrations. To secure your spot at this highly anticipated event, we urge you to complete your registration as soon as possible.\*\*\*\*

Rules for Submission of Camera Ready Manuscript:

1. This email provides you with all the information required to complete your paper and submit it for inclusion in the proceedings. A notification of the timing of the presentation will be sent in a subsequent email.

Here are the steps:

IMPORTANT NOTES:

- 1.1 Please address all the necessary REVIEWERS' COMMENTS (available with your cmt account) which are intended to improve the final manuscript. Final acceptance is conditional on the appropriate response to the requirements and comments. Authors are requested to attach the separate response sheet in pdf along with the revised manuscript. Be careful at the time of mentioning the email id in final manuscript. Email id once finalized will not be changed.

- 1.2 Please prepare your manuscript for final camera-ready submission by following the formats described on the conference web site and using the LNNS Series templates:

<http://maitri.stemrs.in/download.php>

OR

<https://www.springer.com/us/authors-editors/conference-proceedings/conference-proceedings-guidelines>

If your final manuscript does not comply with the formatting requirements and addressing reviewer comments, you will be asked to rectify and resubmit until all formatting requirements are satisfied. Once you are ready to submit, please submit a zip folder that includes the

- Final manuscript (in PDF and MS Word / Latex)
- Copyright (duly signed and scanned)
- Response sheet (Word File with duly signed)
- Ethics (duly signed and scanned)(Download from <http://www.socra.in/downloads/ethics.pdf>)

at [maitri.ic@nitj.ac.in](mailto:maitri.ic@nitj.ac.in) with a subject line: MAITRI2024 PaperID\_no. of paper for e.g. if paper id is 0001

2. Anurag Sharma “Advancements in Hate Speech Detection: A Comprehensive Review of Graph Convolutional Network (GCN) Models”, Accepted at “**International Conference on Artificial Intelligence, Machine Learning and Big Data Engineering (ICAIMLBDE) organized by ISETE**” on 05<sup>th</sup> May 2024 at Hyderabad, India.

Paper id: IST-BDE-HDBD-050524-5541

4/24/24, 11:22 AM

Online Payment



Payment Detail

Payment Confirm

Payment Checkout

Payment Status

|                |  |
|----------------|--|
| Payment Id     | 15194-EXTCONFERENCE2417997   |
| Paper Id       | <b>IST-BDE-HDBD-050524-5541</b>  |
| Name           | Anurag Sharma  |
| Email          | annuragsharma2311@gmail.com  |
| Phone          | 7073862138   |
| Country        | India  |
| State          | Bihar  |
| City           | Darbhanga  |
| Postal Code    | 846003   |
| Address        | Ayachi Nagar, Benta, Po- DMC, opposite Nepali Camp, ward-35, Laheriasarai, Darbhanga |
| Date           | 2024-04-23   |
| Amount         | 7380   |
| Payment Status | Success  |

COPYRIGHT © 2014 ALL RIGHTS RESERVED



IST-BDE-HDBD-050524-5541

## INTERNATIONAL SOCIETY FOR ENGINEERING AND TECHNICAL EDUCATION

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

Organized by: ISETE | Hyderabad, India | 05<sup>th</sup> May 2024


# Certificate of Presentation

This is to certify that *Anurag Sharma* has presented a paper entitled  
*"Advancements in Hate Speech Detection: A Comprehensive Review of  
Graph Convolutional Network (GCN) Models"* at the International  
Conference on Artificial Intelligence, Machine Learning and Big Data  
Engineering (ICAIMLBDE) held in Hyderabad, India

on 05<sup>th</sup> May, 2024.



  
Conference Coordinator  
International Society for Engineering  
and Technical Education

  
Chairman  
International Society for Engineering  
and Technical Education

[www.isete.org](http://www.isete.org)

[info.iseteconference@gmail.com](mailto:info.iseteconference@gmail.com)



## 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 “**Advancements in Hate Speech Detection: A Comprehensive Review of Graph Convolutional Network (GCN) Models**” has been selected for **International Conference on Artificial Intelligence, Machine Learning and Big Data Engineering (ICAIMLBDE)** on **05<sup>th</sup> May 2024 at Hyderabad, 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 Number*) and certificates of Presentation will be given.

### Important Information:

|   |  |
|---|--|
| <b>Paper Title</b>  | <b>Advancements in Hate Speech Detection: A Comprehensive Review of Graph Convolutional Network (GCN) Models</b> |
| <b>Universal paper ID</b><br>(Mention this while Communicating in future) | <b>IST-BDE-HDBD-050524-5541</b>  |
| <b>Author</b>   | <b>Anurag Sharma</b>   |
| <b>Conference link</b>  | <b><a href="https://isete.org/Conference/23969/ICAIMLBDE/">https://isete.org/Conference/23969/ICAIMLBDE/</a></b> |

**NOTE:** Your paper has also cleared the Stage-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 NAME                     | AUTHOR                         |
| MajorProjectReport_cse06.docx  | Anurag Sharma                  |
| WORD COUNT                     | CHARACTER COUNT                |
| 9806 Words                     | 57285 Characters               |
| PAGE COUNT                     | FILE SIZE                      |
| 52 Pages                       | 5.0MB                          |
| SUBMISSION DATE                | REPORT DATE                    |
| May 18, 2024 12:48 PM GMT+5:30 | May 18, 2024 12:50 PM GMT+5:30 |

● 7% Overall Similarity

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



- 4% Internet database
- 3% Publications database
- Crossref database
- Crossref Posted Content database
- 5% Submitted Works database

● Excluded from Similarity Report

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

# Anurag Sharma

## MajorProjectReport\_cse06.docx

 My Files  
 My Files  
 Delhi Technological University

### Document Details

Submission ID  
trnoid::2753559490484

Submission Date  
May 18, 2024, 12:48 PM GMT+5:30

Download Date  
May 18, 2024, 12:52 PM GMT+5:30

File Name  
MajorProjectReport\_cse06.docx

File Size  
5.0 MB

52 Pages  
9,806 Words  
57,285 Characters

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 Anurag Sharma, 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 9.15 CGPA in the first three semesters of my MTech.

I completed my BTech in Computer Science and Engineering from Manipal University Jaipur in 2019. After that, I Joined GAIL India Limited for 1 year as an Apprenticeship Trainee where I worked on several projects 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 500 questions in GFG and LeetCode.

