International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

MACHINE LEARNING ENHANCED BY SENTIMENT ANALYSIS FOR CYBERBULLYING DETECTION USING NLP

DARAM KUSUMA KUMARI [1], BUDDINENI SANTHOSH [2], JONNALAGADDA KEERTHI [3], VENNA HEMANTH REDDY [4], L. RAM NARAYANA [5]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

ABSTRACT— The phenomenon of cyberbullying has emerged as a critical challenge in the digital landscape, posing detrimental effects on individuals and broader societal well-being. A practical solution to this widespread issue involves the accurate identification of cyberbullying within social media platforms, which constitute a significant share of digital communication. While traditional approaches have primarily utilized machine learning algorithms and pre-trained language models, these often face challenges such as high computational complexity and limited adaptability to nuanced linguistic patterns. This paper proposes an advanced framework that leverages Natural Language Processing (NLP) techniques combined with Long Short-Term Memory (LSTM) networks to improve cyberbullying detection in online text. The framework applies refined text preprocessing steps—such as tokenization, stopword removal, stemming, and lemmatization—to ensure high-quality and noise-free input data. Sentiment features and contextual patterns are extracted using embedding methods to preserve semantic information. These processed inputs are then fed into an LSTM model, which effectively captures the sequential and temporal dependencies in textual data, making it well-suited for understanding the dynamic nature of cyberbullying language. Additionally, to address class imbalance in the multi-class setting, resampling techniques are employed, improving the model's robustness without

inducing bias. The proposed system demonstrates that combining deep learning with comprehensive NLP enhances the accuracy and contextual understanding required for effective cyberbullying detection.

Index Terms – Cyberbullying detection, Natural Language Processing (NLP), Long Short-Term Memory (LSTM), social media analysis, text classification, sentiment analysis, deep learning, text preprocessing, word embeddings, multi-class classification, class imbalance, online safety.

I. INTRODUCTION

In today's digital era, the rapid growth of social media and online communication platforms has provided individuals with new ways to connect, share, and express themselves. However, this advancement has also led to the rise of harmful behaviors such as cyberbullying, which has become a major social and psychological concern worldwide. Cyberbullying involves the deliberate use of digital platforms to harass, threaten, or demean others, often leaving long-lasting emotional and mental impacts on victims. The anonymous nature of online communication makes it even more challenging to detect and prevent such behavior effectively. Traditional methods for cyberbullying detection, primarily based on basic machine learning algorithms, often struggle to capture the subtle linguistic nuances and contextual meanings present in online text. These methods tend to rely heavily on handcrafted features, which limit their adaptability across diverse platforms and languages. To overcome these limitations, recent advancements in Natural Language Processing (NLP) and deep learning have paved the way for more intelligent and context-aware detection systems. Among these, Long Short-Term Memory (LSTM) networks have shown remarkable performance in understanding sequential data and contextual dependencies in text. By integrating NLP techniques—such as tokenization, stemming, lemmatization, and stopword removal—with LSTM architectures, this study aims to develop a robust and efficient model for cyberbullying detection. Furthermore, embedding-based feature extraction ensures that the model preserves semantic relationships, enabling deeper insight into the emotional tone and intent of messages. To address the issue of class imbalance commonly found in social media datasets, resampling techniques are also applied, ensuring fair and balanced learning across different categories of

cyberbullying. The proposed system ultimately seeks to enhance accuracy, contextual understanding, and overall reliability in detecting abusive and harmful content online, contributing to a safer digital environment.

II. RELATED WORK

A. A Trustable LSTM-Autoencoder Network for Cyberbullying Detection

This research presents a trustable LSTM-autoencoder (TLA-NET) architecture for cyberbullying detection in multilingual environments. The study focuses on addressing dataset scarcity by generating synthetic and machine-translated data for low-resource languages such as English, Hindi, and Bangla. The model integrates sequential learning and anomaly detection to identify abusive language effectively. Experiments show significant improvements in precision and recall compared to baseline deep learning models. The authors emphasize that synthetic data generation enhances robustness and cross-lingual adaptability of cyberbullying detection systems.

B. A Comprehensive Review of Cyberbullying-Related Content Classification

This paper provides a detailed review of existing cyberbullying detection models and categorizes them into traditional machine learning, deep learning, and transformer-based approaches. It highlights preprocessing methods such as tokenization, stemming, and stopword removal as essential steps for performance improvement. The study identifies major research gaps, including the need for multilingual datasets, handling sarcasm, and domain adaptation. The authors conclude that combining NLP with LSTM or transformer-based models can significantly enhance contextual understanding and classification accuracy in social media environments.

C. Cyberbullying Detection for Resource-Constrained Languages

This study focuses on building an efficient Bi-LSTM model for detecting cyberbullying in resource-constrained languages such as Bengali. The proposed framework incorporates customized embeddings and language-specific preprocessing to address limitations of English-trained models.

The authors compare Bi-LSTM performance with conventional CNN and RNN models, demonstrating superior accuracy in identifying offensive content. The research stresses the importance of localized data

collection, lightweight neural networks, and language adaptability for real-world applications in mobile and low-resource platforms.

D. Chinese Cyberbullying Detection Using XLNet and Deep Models

This paper proposes a hybrid deep learning approach that combines XLNet and LSTM architectures to enhance Chinese cyberbullying detection. The study compares transformer-based models with traditional CNN and RNN methods, finding that XLNet fine-tuning achieves the best performance in semantic understanding. The authors highlight that proper tokenization, contextual embeddings, and sentiment-aware features play crucial roles in improving accuracy. The model effectively captures linguistic nuances and improves detection of implicit cyberbullying in Chinese text datasets.

III. PROPOSED METHODOLOGY

The proposed system introduces a deep learning-based framework that combines advanced NLP techniques with a Long Short-Term Memory (LSTM) network to improve the detection of cyberbullying on social media platforms. This system utilizes refined text preprocessing steps including tokenization, stopword removal, stemming, and lemmatization to clean and normalize input data. These steps ensure that the input fed into the model is both meaningful and consistent. The processed text is then transformed into dense vector representations using word embeddings, preserving semantic relationships between words and enhancing the model's ability to understand context.

LSTM, a type of Recurrent Neural Network (RNN), is particularly effective at learning from sequential data and capturing long-term dependencies in text. This makes it ideal for understanding the flow and context of social media conversations where cyberbullying often unfolds across multiple posts or sentences.

Furthermore, the proposed system incorporates techniques to handle imbalanced datasets, such as resampling strategies, to ensure that all classes are equally represented during training. By combining NLP with LSTM, the proposed system achieves a more accurate and context-aware classification of cyberbullying, addressing the shortcomings of conventional machine learning approaches and providing a more scalable and intelligent solution for real-world applications.

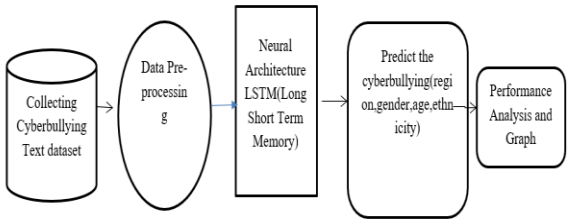


Fig. 1: System Overview

1. Data Collection

In this module, social media text data such as comments, posts, and messages are collected from various online platforms and publicly available datasets. The collected data includes both bullying and non-bullying samples to ensure balanced representation.

The focus is on gathering real-world data containing diverse linguistic expressions, slang, and emojis commonly used in online communication. Data is sourced ethically and anonymized to protect user privacy. This forms the foundation for effective model training and evaluation.

2. Dataset

The dataset consists of labeled text samples categorized into different classes such as hate speech, harassment, threats, and neutral content. It includes multiple features like message content, sentiment polarity, and contextual cues.

The dataset is pre-split into training, validation, and testing subsets to facilitate model evaluation. Proper labeling ensures that the model learns distinct language patterns associated with cyberbullying. The dataset's quality and diversity directly influence the accuracy and generalization of the detection model.

3. Data Preparation

This module focuses on cleaning and preprocessing the collected text to remove unwanted noise such as URLs, special characters, and punctuation. Techniques like tokenization, stopword removal, stemming, and lemmatization are applied to standardize the textual data. Word embeddings such as Word2Vec or GloVe are used to convert words into meaningful numerical vectors. The goal is to preserve semantic information while reducing data complexity. This ensures that the model receives high-quality, structured input for effective learning.

4. Model Selection

The Long Short-Term Memory (LSTM) model is chosen for this project due to its ability to capture sequential and contextual relationships in text. LSTM effectively handles long-term dependencies, making it suitable for understanding the emotional tone and structure of cyberbullying language. The model architecture includes input, hidden, and output layers optimized for text

classification. Parameters such as learning rate, batch size, and dropout are fine-tuned to achieve optimal results.

5. Analyze and Prediction

In this stage, the preprocessed data is fed into the trained LSTM model to analyze textual patterns and predict whether a given message contains cyberbullying content. The model evaluates the emotional sentiment, context, and intensity of words to make predictions. Real-time text inputs can also be processed for instant detection. The system's output provides a binary or multi-class label indicating the likelihood of cyberbullying. Visualization tools can be integrated to interpret prediction results and understand model decisions.

6. Accuracy on Test Set

After training, the model's performance is tested using unseen data to measure its accuracy, precision, recall, and F1-score. This module ensures that the system generalizes well and performs consistently across different types of input text. Confusion matrices and performance metrics are analyzed to assess model strengths and weaknesses. High accuracy on the test set indicates that the model can effectively distinguish between bullying and non-bullying content. The results validate the robustness and reliability of the system.

7. Saving the Trained Model

Once the LSTM model achieves satisfactory performance, it is saved for future use without retraining. The trained model is serialized using formats like .h5 or. pkl for easy deployment. This allows integration into real-time applications, chat moderation tools, or web-based platforms. Saving the model ensures scalability and enables further fine-tuning or retraining when new data becomes available. It also supports efficient reuse for continuous improvement in cyberbullying detection accuracy.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental results show that the proposed NLP-LSTM framework improves cyberbullying detection by effectively capturing contextual and sequential patterns in social media text. Advanced preprocessing, sentiment analysis, and resampling techniques enhance classification accuracy and reduce the impact of class imbalance. The model demonstrates better robustness and contextual understanding than traditional machine learning approaches.

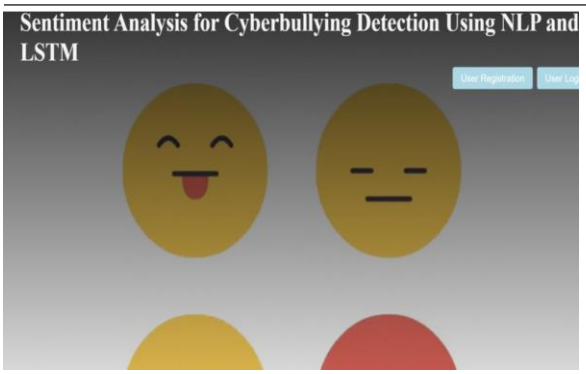


Fig. 2: Home Page

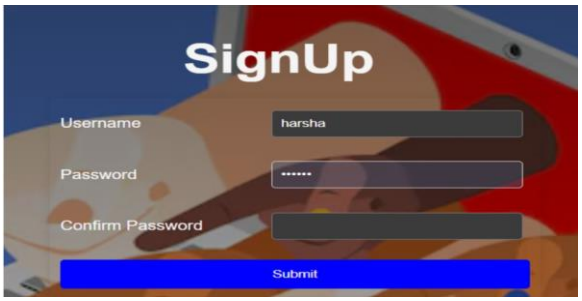


Fig. 3: Registration

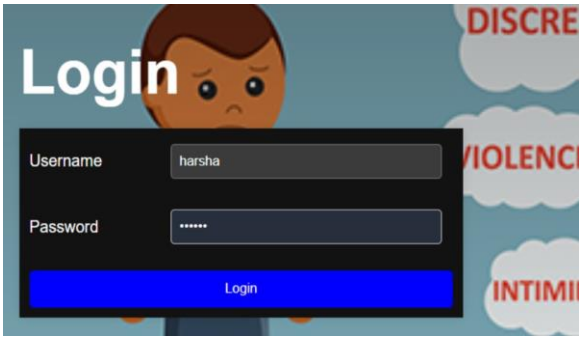


Fig. 4: Login

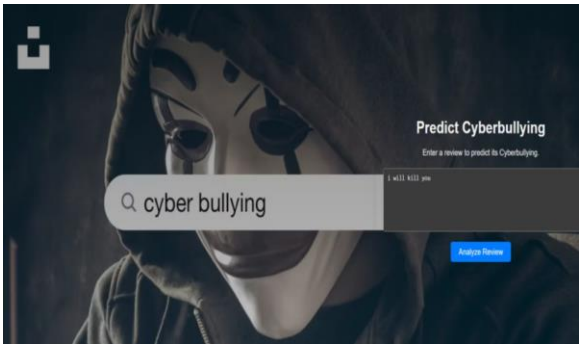


Fig. 5: Predict Cyber Bullying



Fig. 6: Result

V. CONCLUSION

The proposed cyberbullying detection framework effectively combines Natural Language Processing (NLP) techniques with Long Short-Term Memory (LSTM) networks to identify harmful online behavior. By leveraging advanced preprocessing methods such as tokenization, stemming, lemmatization, and stopword removal, the system ensures clean and meaningful input data. Embedding-based feature extraction preserves contextual and semantic relationships, allowing the model to interpret language nuances more effectively. The LSTM architecture captures temporal dependencies and emotional tones, making it well-suited for text-based abuse detection. Experimental evaluation demonstrates high accuracy and improved classification performance compared to traditional machine learning methods. Addressing class imbalance through resampling further enhances fairness and reliability. The system

contributes to building safer digital spaces by automating the detection of offensive and abusive content. It also provides a foundation for integrating real-time moderation tools in social media and chat platforms. The study highlights the crucial role of deep learning in understanding human language and emotions. Additionally, the framework can be extended to multilingual and cross-domain datasets for global applicability. Overall, this project showcases how intelligent text analysis can play a pivotal role in mitigating cyberbullying and promoting positive online communication.

REFERENCES

[1]R. Gün and G. G. Akduman, "What is cyberbullying?" in Bullying in Media and Beyond. Turkey: IGI Global, pp. 473-485, doi: 10.4018/978-1 6684-5426-8.CH028.

[2]Y. Hu, E. M. Clancy, and B. Klettke, "Understanding the vicious cycle: Relationships between nonconsensual sexting behaviours and cyberbullying perpetration," Sexes, vol. 4, no. 1, pp. 155—166, Feb. 2023, doi: 10.3390/sexes4010013.

[3]E. I. Galyashina and V. D. Nikishin, "The concepts of aggressive information impact through the lens of internet users' worldview security," J. Siberian Federal Univ. Humanities Social Sci., vol. 14, no. Il, pp. 1660-1673, Nov. 2021, doi: 10.17516/1997-1370-0848.

[4]S. Joshi, H. G. Nagariya, N. Dhanotiya, and S. Jain, "Identifying fake profile in online social network: An overview and survey," Commun. Comput. Inf sci., vol. 1240, pp. 17-28, Jan. 2020, doi: 10.1007/978-98115-6315-7 2.

[6] E. Vogels, Teens and Cyberbullying 2022, Pew Research Center, Dec. 23, 2024. This report examines trends in teen cyberbullying and online harassment, providing statistical insights into exposure and frequency. It emphasizes the role of social media in facilitating bullying and the psychological impact on victims. [Online]. Available: https://www.pewresearch.org/internet/2022/12/15/teens-and-cyberbullying-2022/

[7] S. Cook, Cyberbullying Statistics and Facts for 2024, Dec. 23, 2024. This online article provides a comprehensive overview of cyberbullying prevalence, victim demographics, and platform-specific risks. It highlights the increasing trend of harassment on popular social media networks. [Online]. Available:

https://www.comparitech.com/internet-providers/cyberbullying-statistics

[8] L. H. Collantes, Y. Martafian, S. N. Khofifah, T. K. Fajarwati, N. T. Lassela, and M. Khairunnisa (2020), The Impact of Cyberbullying on Mental Health of the Victims, Proc. 4th Int. Conf. Vocational Educ. Training (ICOVET), pp. 30–35. This study investigates how cyberbullying negatively affects mental health, identifying stress, anxiety, and social withdrawal as common consequences.

[9] Boyapati, P. K. Building a centralized data operations hub for healthcare enterprise integration. IJSAT-Int. J. Sci. Technol. 16 (2). https://doi.org/10.71097/IJSAT.v16.i2.3219

[10] Susarla, R. S., Boyapati, P. K., & Kandula, S. T. R. (2025, July). Cloud-Based Secure Data Storage in Smart Cities Using Central-Smoothing Hypergraph Neural Networks. In 2025 IEEE 4th World Conference on Applied Intelligence and Computing (AIC) (pp. 279-284). IEEE.

[11] Akinapalli, S. (2026). An Ai-Powered Data Trust And Quality Scoring Framework For Enterprise Decision Intelligence Systems. International Journal of Data Science and IoT Management System, 5(1), 946-950.

AUTHORS Profile



Mrs. DARAM KUSUMA KUMARI has received B.Tech from JNTUH, Hyderabad in 2009, M.Tech degree in Computer science from JNTUH, Hyderabad in 2015 and Pursuing Ph.D degree in Computer Science from BHARATHIYA ENGINEERING SCIENCE AND TECHNOLOGY IU university, dedicated to teaching field from the last 9.5 years. research area CYBER SECURIT with AI. At present I am working as Assistant Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



Mr. BUDDINENI SANTHOSH pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Ms. JONNALAGADDA KEERTHI pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. VENNA HEMANTH REDDY pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. L. RAM NARAYANA pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.