IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

AI Based Multilingual Cyberbullying Detection And Auto Comment Deletion System

Mrs. A. Rubika, S. Akila, G. Sanjana, M. Poonkodi, B. Renuka Devi
Department of Computer Science and Technology
Vivekanandha College of Engineering for Women (Autonomous) Tiruchengode, Namakkal, Tamil Nadu, India
akilasuresh3092005@gmail.com , rubikaanbu1988@gmail.com , sanjuganesan006@gmail.com ,
mpoonkodi45@gmail.com renukadevivpt@gmail.com

Abstract: *The rapid growth of social media platforms has significantly increased user interaction, but it has also led to a rise in cyberbullying and abusive communication. This paper presents an AI-Based Multilingual Cyberbullying Detection and Auto Comment Deletion System designed to identify and remove harmful comments in real-time. The proposed system is implemented as a web-based social media application similar to Instagram, where users can create profiles, post content, and interact through comments. The system utilizes Natural Language Processing (NLP) and Machine Learning techniques to detect abusive and cyberbullying content across multiple languages. When a user posts a comment, the system analyzes the text instantly and automatically deletes it if it contains offensive or harmful content. This approach ensures a safer online environment by preventing the spread of toxic language. The system also supports multi-user interaction and real-time processing, making it scalable and efficient. The proposed solution contributes to creating a positive and secure digital communication space*

Keywords: Cyberbullying Detection, NLP, Machine Learning, Multilingual System, Auto Comment Deletion, Social Media Analysis

I. INTRODUCTION

Social media platforms have become an essential part of daily communication, enabling users to share ideas, images, and opinions. However, the increasing usage of these platforms has also resulted in the growth of cyberbullying, harassment, and abusive language. These negative interactions can severely impact users' mental health and online experience. AI-based multilingual cyberbullying detection system integrated into a web application. The system is designed to function like a social media platform, allowing users to create profiles, post content, and interact with others through comments. The key objective of the system is to automatically detect and delete abusive comments in real-time. By using advanced machine learning algorithms and natural language processing techniques, the system can identify harmful content across multiple languages. This ensures that users are protected from cyberbullying regardless of the language used.

II. LITERATURE SURVEY

Several research studies have been conducted on cyberbullying detection using machine learning and deep learning techniques.

[1] K. Dinakar et al. (2011) proposed a machine learning-based approach for detecting cyberbullying in textual data. Their work focused on using classification algorithms to identify bullying-related content in social media conversations. The study highlighted the importance of context-aware detection for improving accuracy.

[2] J. Xu et al. (2012) developed a system to detect bullying traces in online platforms using NLP techniques. The research emphasized analyzing user comments and conversations to identify patterns of abusive behavior, which helped in early detection of cyberbullying.

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 2, April 2026 Impact Factor: 8.2

[3] A. Schmidt and M. Wiegand (2017) presented a comprehensive survey on hate speech detection. Their work discussed various machine learning approaches such as Support Vector Machines (SVM), Naive Bayes, and deep learning models. The study concluded that combining linguistic features with machine learning improves detection performance.

[4] Y. Chen et al. (2019) proposed a deep learning-based approach using Long Short-Term Memory (LSTM) networks to detect offensive language. The model showed improved performance in capturing sequential patterns in text, making it more effective than traditional methods.

[5] T. Davidson et al. (2017) focused on distinguishing between hate speech and offensive language using supervised learning techniques. Their work addressed the challenge of misclassification and improved the precision of detection systems.

[6] J. Devlin et al. (2018) introduced BERT (Bidirectional Encoder Representations from Transformers), a powerful language model that significantly improved text classification tasks. BERT-based models have been widely used for detecting abusive language with high accuracy.

Recent studies have focused on multilingual cyberbullying detection, where systems are designed to detect abusive content in multiple languages. These approaches use translation techniques and multilingual embeddings to ensure effective detection across diverse user inputs.

III. METHODOLOGY

3.1 User Management Module : This module is responsible for handling user registration, login, and authentication processes within the system. New users can create accounts by providing necessary details such as username, email, and password, while existing users can securely log in using their credentials. The module ensures data privacy and security by validating user inputs and preventing unauthorized access. It also manages user profiles, allowing users to update their information and maintain their activity history. This module plays a crucial role in enabling personalized user experience and maintaining system integrity.

3.2 Post and Interaction Module : This module enables users to create, upload, and share posts similar to popular social media platforms like Instagram. Users can upload images along with captions and view posts shared by other users in the system. Additionally, this module supports user interaction through comments, allowing users to express their opinions or feedback on posts. The interaction system is designed to handle multiple users simultaneously, ensuring smooth communication. This module increases user engagement and acts as the primary interface for communication within the platform.

3.3 Text Preprocessing Module : This module prepares user-generated comments for analysis by the machine learning model. Raw text data often contains noise such as special characters, emojis, and irrelevant words, which can affect model performance. Therefore, preprocessing steps such as text cleaning, tokenization, stop-word removal, stemming, and normalization are applied. These techniques convert the text into a structured format that can be easily processed by the detection model. By improving the quality of input data, this module significantly enhances the accuracy and efficiency of cyberbullying detection.

3.4 Cyberbullying Detection Module : This is the core module of the system that identifies abusive or harmful content in user comments. It uses machine learning algorithms combined with Natural Language Processing (NLP) techniques to analyze the textual data. The model is trained using labeled datasets containing both abusive and non-abusive comments, enabling it to classify new inputs effectively. The module supports multilingual detection, allowing it to identify cyberbullying across different languages. It continuously evaluates the comment content and provides a classification result, which is then used by the deletion module for further action.

3.5 Auto Comment Deletion Module : This module is responsible for automatically removing comments that are identified as abusive by the detection system. Once a comment is classified as harmful, it is immediately deleted before being displayed to other users. This real-time action prevents the spread of offensive content and ensures a safer online

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

environment. The module operates efficiently without manual intervention, reducing the workload of administrators. It plays a key role in maintaining platform integrity and promoting positive user interaction.

3.6 Auto Comment Deletion Module : This module is responsible for automatically removing comments that are identified as abusive by the detection system. Once a comment is classified as harmful, it is immediately deleted before being displayed to other users. This real-time action prevents the spread of offensive content and ensures a safer online environment. The module operates efficiently without manual intervention, reducing the workload of administrators. It plays a key role in maintaining platform integrity and promoting positive user interaction.

IV. SYSTEM ARCHITECTURE

4.1 User Interface Module : This module acts as the front-end of the system where users can register, log in, and access all features of the application. It allows users to create profiles, upload posts, and interact with other users through comments. The interface is designed to be user-friendly and responsive, ensuring smooth navigation. It serves as the main communication point between the user and the system.

4.2 Text Preprocessing Module : This module is responsible for cleaning and preparing user-generated comments before analysis. It removes unwanted elements such as special characters, emojis, and stop words, and performs tokenization and normalization. These steps convert raw text into a structured format suitable for machine learning models. Proper preprocessing helps improve the accuracy and efficiency of the detection system.

4.3 Cyberbullying Detection Module : This module is the core component that analyzes comments using machine learning and NLP techniques. It is trained on datasets containing both abusive and non-abusive text to classify user inputs effectively. The module supports multilingual detection, allowing it to handle comments in different languages. It produces a classification result that determines whether the content is safe or harmful.

4.4 Auto Comment Deletion Module : This module automatically removes comments that are identified as abusive by the detection system. The deletion process occurs in real-time, preventing harmful content from being displayed to other users. This reduces the spread of cyberbullying and maintains a positive environment. The module works without manual intervention, ensuring efficiency and consistency.

4.5 Database Module : This module manages the storage of all system data, including user information, posts, and comments. It ensures secure data handling and supports efficient retrieval when required. The database is designed to handle multiple users and large volumes of data simultaneously. It plays a vital role in maintaining system performance and reliability.

4.5 FLOWCHART

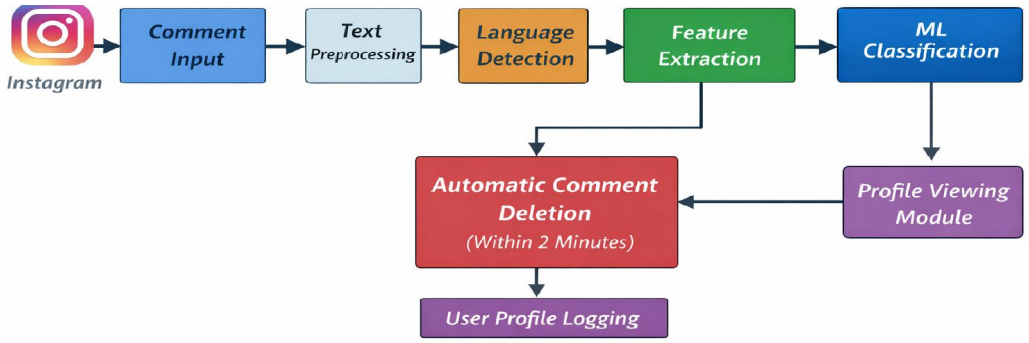

FIGURE 4.1 AI Based Cyberbullying Detection And Auto Comment Deletion System

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

V. RESULTS AND DISCUSSION

5.1 OUTPUT

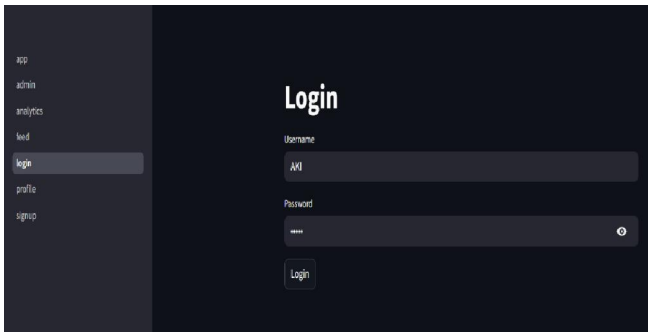

FIGURE 5.1

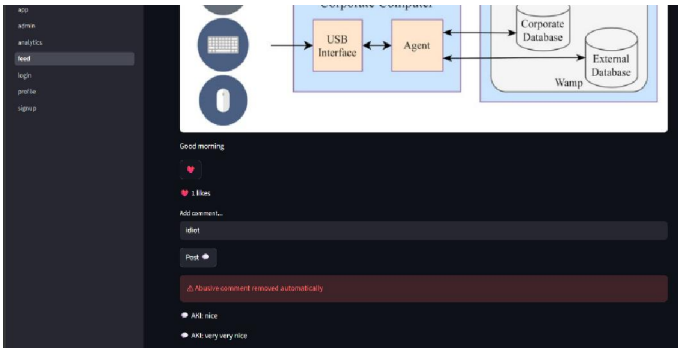

FIGURE 5.2

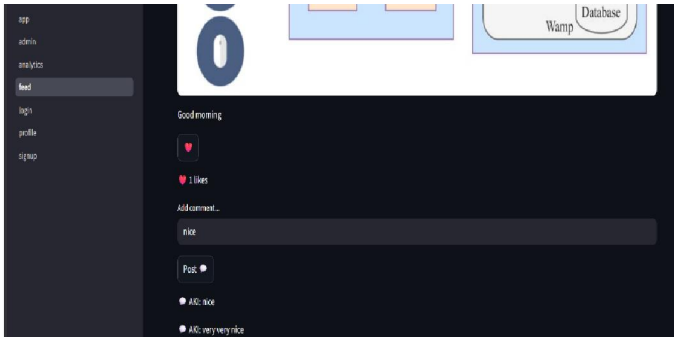

FIGURE 5.3

5.2 CHALLENGES AND SOLUTIONS : The development of this system faces challenges such as limited availability of multilingual datasets and variations in slang or informal language used in social media comments. Real-time detection and accurate classification of abusive content are also difficult due to processing constraints and possible misclassification. To overcome these issues, data preprocessing and data augmentation techniques are applied to improve data quality. Efficient machine learning models and optimization methods are used to ensure accurate detection and fast real-time performance.

V. CONCLUSION

This paper presents an AI-Based Multilingual Cyberbullying Detection and Auto Comment Deletion System designed to identify and remove abusive comments in real-time. By integrating machine learning and Natural Language

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 2, April 2026 Impact Factor: 8.2

Processing techniques, the system effectively classifies and filters harmful content across multiple languages. The proposed solution enhances user safety and promotes a positive online environment. Overall, the system provides an efficient and scalable approach to prevent cyberbullying on social media platforms.

REFERENCES

[1] T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated Hate Speech Detection and the Problem of Offensive Language," Proceedings of ICWSM, 2017. (widely used baseline work)
[2] Y. Chen, Y. Zhou, S. Zhu, and H. Xu, "Detecting Offensive Language in Social Media using Deep Learning," IEEE Access, 2019. (deep learning approach)
[3] A. Al- Garadi et al., "Deep Learning-Based Cyberbullying Detection in Social Media," Future Generation Computer Systems, 2020.
[4] S. R. Jha and V. Mamidi, "When Does a Compliment Become Sexist? Analysis and Classification of Ambiguous Comments," Proceedings of ACL, 2020.
[5] M. Mozafari, R. Farahbakhsh, and N. Crespi, "Hate Speech Detection and Racial Bias Mitigation in Social Media Based on BERT Model," IEEE Access, 2020.
[6] Z. Waseem, T. Davidson, D. Warmsley, and I. Weber, "Understanding Abuse: A Typology of Abusive Language Detection," 2021.
[7] S. Vidgen et al., "Learning from the Worst: Dynamically Generated Datasets for Hate Speech Detection," ACL, 2021.
[8] Y. Chen, Y. Zhou, S. Zhu, and H. Xu, "Detecting Offensive Language in Social Media using Deep Learning," IEEE Access, 2019.