Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Cyberbullying detection and prevention in webchat applications using SVM and CNN

Dr.R.V.Patil[1], Rohan Chaudari[2], Gayatri Jadhav[3], Rutuja Choudhary[4], Sushant Aivale[5]
Department of Computer Engineering
PDEA'S College of Engineering, Manjari
Pune

Abstract -: As we all know in today's digital era with the rapid growth of online communication, web chat applications have become an essential part of our day-to- life. But however, this growth has also led to a significant and considerable rise in cyberbullying, where individuals use digital platforms to harass, threaten, or harm other individuals . Due to the fast-growing and often anonymous nature of chat environments, identifying and controlling such behavior has become increasingly difficult and challenging . Our paper presents an effective approach to detect and prevent cyberbullying in web-based chat systems using a combination of machine learning and deep learning techniques, specifically Support Vector Machines (SVM) and Convolutional Neural Networks (CNN).

The proposed system processes chat messages using techniques such as tokenization, and word representation. SVM is used to classify messages based on learned patterns from structured data, while CNN helps in capturing deeper contextual meaning and relationships within the text. By combining these two approaches, the system is able to achieve more accurate and reliable detection compared to using a single model.

The framework is designed as a step-by-step pipeline that includes data collection, preprocessing, feature extraction, model training, and real-time classification. It is tested on standard datasets containing both harmful and non-harmful messages. The results show strong performance across key evaluation metrics such as accuracy, precision, recall, and F1-score, demonstrating the effectiveness of the combined model.

Beyond detection, the system also focuses on prevention by introducing features such as real-time warnings, message filtering, and monitoring of user behavior. These measures aim to reduce the spread of harmful content, protect one's mental peace and encourage safer and non-bullying communication practices.

Overall, this work provides a practical and scalable solution for improving online safety in chat applications. It can be applied to social media platforms, messaging services, and educational tools to help create a more respectful and secure digital environment.

I. INTRODUCTION

In recent years, the way people interact has changed completely with the great use of the internet and web-based chat applications. Platforms such as messaging apps, online sites, and social networking sites have made communication faster, easier, and more accessible than ever before. While these technologies have brought people closer and created new opportunities for interaction, they have also given rise to serious and considerable issues—one of the most viewed is being **cyberbullying.** Cyberbullying refers to the use of digital platforms to harass, threaten, or target individuals through harmful messages, abusive language, or repeated negative behavior inorder to destroy any individual . Unlike traditional bullying, cyberbullying can occur at any time and reach a wider group of people instantly. The presence of anonyomous nature in many web chat applications often encourages users to follow in such behavior without fear of immediate results . This not only affects the mental and emotional well-being of an individual but can also lead to long-term psychological impacts such as anxiety, depression, and social withdrawal etc.

To address these issues , this paper looks forward to the use of machine learning and deep learning techniques for cyberbullying detection and prevention. In particular, it focuses on two widely used approaches: **Support Vector Machines (SVM)** and **Convolutional Neural Networks (CNN)**. SVM is known for its effectiveness in handling high-dimensional text data and providing reliable classification results, especially when the dataset is well-structured. On the other hand, CNN, a deep learning model, capable of automatically extracting important features from text and understanding contextual relationships between words. By combining these two approaches, the system aims to strengthen both models to improve detection accuracy in chats.

The developed system follows a structured workflow that begins with data collection from chat-based datasets, followed by preprocessing steps such as , tokenization, and feature extraction. The processed data is then used to train both SVM and CNN models. Once the model is trained, the system can classify incoming chat messages as either bullying or non-bullying in real time. In addition to detection, the system also includes a prevention mechanism, which can

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

generate alerts, filter harmful messages, and monitor user behavior to reduce the display of cyberbullying.

II. LITERATURE SURVEY

1. Cyberbullying Detection on Social Networks using ML.M islam et al. (2021) tried to design and evaluate a practical framework for automatically detecting cyberbullying in social-media text by combining standard NLP preprocessing with multiple ML classifiers and a character-level CNN variant. The paper compared many classical classifiers (Naïve Bayes, Logistic Regression, SVM, Random Forest) on two social datasets, namely Facebook comments and a Twitter/ Kaggle tweets set, and showed a character-level CNN with shortcuts, Char-CNNS, along with focal-loss in order to better handle spelling noise , confusement, and class imbalance.

2. Review of ML Techniques in Cyberbullying Detection

Sultan et al. conducted a systematic review of the literature to summarize the machine-learning and NLP approaches for automatic detection of cyberbullying on social media. The review focused on the full workflow , from data collection to preprocessing, feature extraction or selection, to machine learning models, and then evaluation. A total of 13 primary studies were drawn from four reference databases: ScienceDirect, IEEE Xplore, Springer, and Wiley, published between 2015 and 2021.

3. ML vs Transfer Learning for Cyberbullying Detection

Teng & Varathan (2023) aimed to take into consideration the results and compared , by means of experiments, the performance of conventional machine-learning-based approaches (TF-IDF / features + classical classifiers) with transfer-learning-based methods (pre-trained transformer encoders and fine-tuning) for text-based cyberbullying detection on social networks. Their main contribution worked from head-to-head experimental comparison on several datasets and modelling workflows, together with an careful analysis of metrics (accuracy, robustness, runtime/complexity) to guide users who have to choose between lightweight Machine Learning and heavier Tansfer Learning solutions.

4. Cyberbullying Detection on Social Media using Supervised ML

Perera & Fernando in 2024 presented an applied system and study that conducted by taking into consideration the observation noted down that evaluates supervised machine learning methods for detecting cyberbullying on social media. The authors have very practical objectives: to create a detection workflow (collection → preprocessing → feature extraction → classifier) and compare various supervised classifiers on social-media text (tweets/comments), then discussesd the deployment of a working cyberbullying detection system.

III. PROBLEM STATEMENT

The rapid growth of web-based chat applications and online communication platforms has made digital interaction an important part of everyday life. However, along with these advancements, there has been a significant increase in number of cyberbullying activities , where individuals use chat platforms to send harmful, abusive, or threatening messages. This issue has become a major concern, especially among young users, as it negatively affects mental health, emotional well-being, and overall online safety. One of the key challenges in addressing cyberbullying is the difficulty of detecting harmful content in real time**.** Chat messages are typically short, informal, and highly unstructured. Users often rely on slang, abbreviations, emojis, misspellings, and coded language to express themselves. In many cases, offensive intent is not directly visible through specific keywords, making simple filtering techniques ineffective. For example, a message may appear normal on the surface but carry a harmful or sarcastic meaning depending on the context. This makes it difficult for traditional rule-based or keyword-based systems to accurately identify cyberbullying.

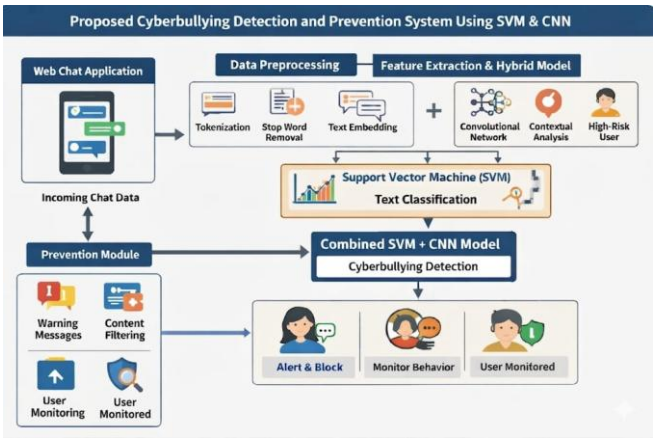
To address these challenges, there is a need for an intelligent and scalable solution that can:

- Accurately detect cyberbullying in short and unstructured chat messages
- Understand contextual and semantic meaning rather than relying only on keywords
- Minimize false positives and false negatives
- Operate in real time for immediate detection
- Provide preventive mechanisms to reduce harmful interactions
- Scale efficiently to handle large volumes of chat data

This paper proposes a hybrid approach that combines Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) to overcome these limitations. By integrating the strengths of traditional machine learning and deep learning techniques, the system aims to improve the accuracy, reliability, and real-time performance of cyberbullying detection while also incorporating preventive measures to enhance user safety in web chat environments**.**

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

IV. PROPOSED SYSTEM



The process begins with incoming chat data from the users interacting on a web chat platform. Since raw text data is often noisy and unstructured, it first undergoes data preprocessing**,** where unnecessary elements are removed and the text is converted into a suitable format. This includes steps such as tokenization (breaking text into words), stop-word removal (eliminating common but irrelevant words), and text embedding (converting text into numerical vectors).After preprocessing, the system performs feature extraction and analysis. Here, the CNN model plays a crucial role by capturing the contextual and semantic meaning of the text. Unlike traditional methods, CNN can understand patterns, tone, and relationships between words, helping to identify subtle forms of cyberbullying. At the same time, features derived from the text are passed to the SVM model

Working (Step-by-Step)

- **1. Input Stage**
  - Chat messages are received from the web chat application as incoming data.
- **2. Data Preprocessing**
  - Text is cleaned and prepared using:
    - Tokenization
    - Stop-word removal
    - Text embedding
- **3. Feature Extraction & Analysis**
  - CNN captures contextual and semantic meaning of messages.
  - Identifies patterns and high-risk behavior.
- **4. Classification (SVM)**
  - SVM performs text classification (bullying vs non-bullying).
- **5. Hybrid Model (SVM + CNN)**
  - Combines outputs to improve accuracy and reliability of detection.
- **6. Detection Output**
  - System labels messages as cyberbullying or safe**.**
- **7. Prevention Module**
  - Takes real-time actions:
    - Sends warning messages
    - Performs content filtering
    - Tracks user monitoring
    - Marks user monitored status
- **8. Final Actions**
  - Alert & block harmful users
  - Monitor behavior continuously
  - Promote safe communication environment

V. PROCESS FLOW

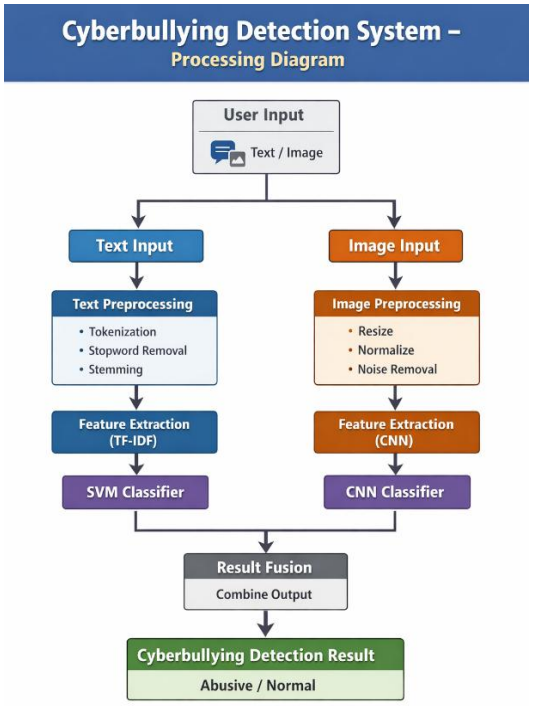


fig. Process flow diagram.

VI. HARDWARE COMPONENTS

Processor (CPU)

- Minimum: Intel i5 / AMD Ryzen 5
- Recommended: Intel i7 / AMD Ryzen 7 or higher
- Reason: Required for preprocessing, feature extraction, and SVM training

RAM

- Minimum: 8 GB
- Recommended: 16–32 GB
- Reason: Handles large text datasets and model training efficiently

Storage

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

- Minimum: 256 GB SSD
- Recommended: 512 GB – 1 TB SSD
- Reason: Faster read/write for datasets, embeddings, and model files

System Type

- 64-bit architecture (Windows/Linux/Mac)
- Preferred: Linux (Ubuntu) for better ML compatibility

VII. CONCLUSION

In today's digital world, web chat applications have become an essential part of everyday communication, but they also create opportunities for harmful behaviors such as cyberbullying. Addressing this issue is not just a technical challenge, but also a social necessity. This project presents a practical and intelligent solution by combining **machine learning (SVM)** and **deep learning (CNN)** to detect and prevent cyberbullying in real time and protect peoples mindset from getting destroyed.

VIII. REFERENCES

[1] J. Hani et al., Social Media Cyberbullying Detection using ML, IJACSA, 2019.

[2] A. Muneer, S. Fati, Comparative Analysis of ML Techniques for Cyberbullying on Twitter, IEEE DSAA, 2020.

[3] M. Islam et al., Cyberbullying Detection on Social Networks using ML, IEEE CSDE, 2021.

[4] S. Neelakandan et al., Deep Learning Approaches for Cyberbullying Detection, Comput. Intell. Neurosci., 2022.

[5] D. Sultan et al., Review of ML Techniques in Cyberbullying Detection, IEEE Access, 2022.

[6] T. H. Teng, K. D. Varathan, Cyberbullying Detection: ML vs Transfer Learning, IEEE, 2023.

[7] A. Almomani et al., Image Cyberbullying Detection using Transfer DL, IJ Cognitive Computing in Eng., 2024.

[8] A. Perera, P. Fernando, Cyberbullying Detection using Supervised ML, Procedia Comput. Sci., 2024.