Journal of Science and Technology Excellence
ISSN: **3108-3102**| Volume 2 Issue 1 January-March 2026
www.jstejournal.com **DOI**: *https://doi.org/*10.5281/zenodo.19147434 **JSTE-07**

CYBERBULLYING DETECTION IN SOCIAL NETWORKS USING MULTIMODAL TECHNIQUES

[#1]J. SWATHI, ***Associate Professor & HOD,***
[#2]A. ARADHANA, ***B.Tech Student,***
[#3]GATTU MEGHANA, ***B.Tech Student,***
[#4]JADI AKANKSHA, ***B.Tech Student,***
[#5]BUDIDHA SHIRISHA, ***B.Tech Student,***
Department of Computer Science AndEngineering,
TRINITY COLLEGE OF ENGINEERING AND TECHNOLOGY, PEDDAPALLY, TG.

ABSTRACT: The issue of cyberbullying on social media is on the rise, and it is imperative that we develop more effective methods to identify it. To improve results, a multi-modal method combines text, pictures, and user behavior. Computer vision and natural language processing are both capable of detecting incorrect text and images. Machine learning systems can detect patterns indicative of cyberbullying by analyzing user connections. This approach improves detection accuracy by considering a wide range of materials. By absorbing data from their environments, deep learning systems enhance categorization accuracy. Stopping online harassment is a breeze with real-time tracking. In order to put an end to various forms of cyberbullying using this approach, continuous instruction is required. With this strategy, internet users will be better protected. In order to improve detection, future research should make use of state-of-the-art AI systems.

KEYWORDS:Cyberbullying, Social Networks, Multi-Modal Approach, Machine Learning, Natural Language Processing (NLP), Computer Vision, Deep Learning, User Behavior Analysis, Real-Time Monitoring, Online Safety.

1. INTRODUCTION

Online networking and information sharing are common pastimes among teenagers. Through social networks, users are able to instantaneously communicate and share information with a large number of people. More than three billion people across the world use social media. Intentionally harming or embarrassing another person online through mobile devices, video game apps, or any other way of conveying text, photos, or videos is called cyberbullying, according to the National Crime Security Council (NCPC). Because the internet is always available, cyberbullying can happen whenever someone wants it to. The use of anonymous text, images, or videos can be a part of cyberbullying. Finding the work's original creator could be an arduous, if not impossible, undertaking. Furthermore, doing away with these transactions altogether in the future is simply not practical. Social media sites including Instagram, Facebook, Snapchat, Skype, and Twitter are prime locations for cyberbullying. Recommendations for avoiding abuse are available on Facebook and other social media platforms. In addition to discussing user blocking, this section details how to report instances of cyberbullying.

Journal of Science & Technology Excellence
Journal of Science and Technology Excellence
ISSN: **3108-3102**| Volume 2 Issue 1 January-March 2026
www.jstejournal.com **DOI**: *https://doi.org/*10.5281/zenodo.19147434 **JSTE-07**

Instagram users can choose to follow accounts they find offensive or unfollow accounts they find harmful. Anyone in the community can report violations and provide ideas for how to make the app better. Aditya Desai can be reached at this email address: adityadesai1703@gmail.com. Investigation and documentation of cyberbullying's past are crucial because the social dynamics of this phenomenon extend beyond the physical boundaries of human connection and include unregulated contacts with strangers. Because it only takes a few clicks of the mouse to engage in cyberbullying, the victim feels violated all the time. There could be physical, mental, and emotional effects on the sufferer. Right present, the most prevalent kind of cyberbullying is the spread of nasty or harassing comments on social media. In order for a system to react correctly, it needs to be able to distinguish between messages that constitute bullying and those that do not.

Social media and messaging apps can help reduce the frequency and impact of cyberbullying with a comprehensive system that detects such instances. Recognizing cyberbullying text and determining its relevance are the main goals of the cyberbullying detection system. One must read a document in its whole to determine its context before using the previously supplied information or images. Efficient and effective access to the document requires the implementation of a suitable method.

2. LITERATURE SURVEY

Tabassum, I., &Nunavath, V. (2024)This work proposes hybrid deep learning models that leverage social media visual and verbal data to detect cyberbullying. Researchers extracted text using LSTM, GRU, BERT, DistilBERT, and RoBERTa. They extracted images using CNN, ViT, and ResNet. To integrate modalities, late fusion was used. RoBERTa+ViT performed multi-class cyberbullying classification with high accuracy and F1-scores on public and private datasets.

Muneer, A., Alwadain, A., Ragab, M.G., &Alqushaibi, A. (2024)The authors describe a deep neural network method for Twitter cyberbullying detection using ensemble stacking learning and a modified BERT model (BERT-M). The study used word2vec with Continuous Bag of Words for feature extraction. Convolutional and pooling approaches lowered dimensionality and captured abusive words. On Twitter, the suggested stacking model outperformed baseline BERT and current NLP detectors with an F1-score of 0.964 and 97.4% accuracy.

Aliyeva, Ç.O., &Yağanoğlu, M. (2024)This study examines Twitter cyberbullying detection using deep learning. This project uses NLP to identify hazardous tweets. Research showed that deep learning can stop social media cyberbullying by improving detection with complicated neural network architectures.

Bhaskar, K., Reddy, S.A., Yatheendra, K., &Prathyusha, G. (2024)Traditional machine learning and transfer learning are used to identify social media cyberbullying. Researchers evaluated Electra, DistilBERT, DistilRoBerta, LinearSVC, and Logistic Regression using feature extraction and selection methodologies.We believed these techniques would help models understand cyberbullying and improve detection systems.

ISSN: **3108-3102**| Volume 2 Issue 1 January-March 2026
www.jstejournal.com **DOI**: *https://doi.org/*10.5281/zenodo.19147434 **JSTE-07**

Ramakrishna, B., Sathvika, B., Chandana, G., &Nandagiri, K. (2024)This analysis identifies social media cyberbullying using a Semantic-enhanced Marginalized Stacked Denoising Autoencoder. The model finds bullying content's key feature structures using sparsity restrictions and semantic dropout noise. SmSDA reduced cyberbullying better than baseline treatments in Twitter and MySpace experiments.

Abood, M.M., & Al-Bayati, M.A. (2024) This study introduces EMDL-CBD, a deep learning-based multimodal cyberbullying detection algorithm. This paradigm considers visual and textual data. Grad-CAM shows the model's picture predictions, while LRP shows its text predictions. After data fusion, accuracy metrics assessed the model's cyberbullying detection ability.

Hegde, C.L. (2024)Information from numerous social media sites is used in this master's thesis to identify and prevent cyberbullying. The system uses machine learning and NLP to detect cyberbullying before victims are hurt. The article's deep dive into machine learning algorithms and feature engineering yielded 87.2% to 95.5% results. Thus, it makes internet users safer.

Kumar, A., & Tripathi, A.K. (2024)This project focuses on improving cyberbullying detection systems using ensemble learning. This study uses multiple machine learning methods to improve detection accuracy and reliability to protect teens' online safety. The ensemble technique has improved cyberbullying detection to improve online safety.

Kumari, K., & Singh, J.P. (2023)This study examines the difficulties of detecting cyberbullying in mixed-media social media posts. The authors merge text and image modalities using a convolutional neural network for text analysis and a pre-trained VGG-16 network for image analysis. These parameters are refined by a genetic algorithm to improve detection. A dataset with text and visuals yielded an F1-score of 78% for the suggested model. This is 9% better than previous findings on the same dataset.

Qiu, J., Moh, M., & Moh, T.S. (2023)This research proposes a multimodal Twitter cyberbullying detection approach using visual and verbal characteristics. The system uses a Tensor Fusion Network, a CNN for textual analysis, and a VGG-19 network for visual analysis. After testing and training on Twitter datasets, the system obtained 93% accuracy, 4% better than benchmark text-only models and 6.6% better than multi-modal algorithms. The findings suggest that incorporating many data sources can improve cyberbullying detection.

Kumar, A., & Sachdeva, N. (2022)The authors present CapsNet-ConvNet, a deep neural model that detects cyberbullying in uploaded photos, essays, and infographics. The design uses convolutional neural networks for image and dynamic routing capsule networks for word analysis. Graphical data analysis with Google Lens separates text from photos. A perceptron-based decision-level late fusion technique integrates multimodal forecasts. The model detected multi-modal cyberbullying with an AUC-ROC of 0.98 using 10,000 YouTube, Instagram, and Twitter posts.

Saichandana, B., & Kamakshi, P. (2022)This investigation explores leveraging textual and aural data from social media sites to identify cyberbullying. To detect textual cyberbullying,

ISSN: **3108-3102**| Volume 2 Issue 1 January-March 2026
www.jstejournal.com **DOI**: *https://doi.org/*10.5281/zenodo.19147434 **JSTE-07**

the authors employed logistic regression, support vector machines, and naive Bayes classifiers. Model performance was assessed using F1-score, recall, accuracy, and precision. The study suggests adding a variety of data categories in online social bullying detection systems.

Nlerum, P.A., &Brisibe, B. (2021)With cyberbullying on the rise, this study provides a hybrid text detection approach using sentiment analysis, unsupervised learning, and LSTM networks. The F1-score, precision, recall, and accuracy were utilized to evaluate the model using the Russian Troll dataset. After employing TF-IDF for feature extraction, tokenization, and sequential modeling with LSTM layers, we achieved 94.9% accuracy and above 95% recall, F1-score, and precision. Our results suggest that the model can handle complex and perplexing social media data.

Hegde, C.L. (2020)This master's thesis proposes a proactive detection and preventive strategy for cyberbullying using data from numerous social media platforms. Machine learning and NLP can detect cyberbullying before it harms targets. The study evaluates machine learning algorithms and feature engineering methods with 87.2% to 95.5% accuracy. Thus, it makes internet users safer.

3 PROPOSED METHODOLOGY

The media should examine mood, syntactic, semantic, and ironic factors when identifying hate speech. The time-tested method of sentiment analysis is used to identify and extract subjective information needed to understand the subject's emotional state, point of view, or attitude from texts in their original context. Our "social" components can detect cyberbullying. All retrievable features were classified into five categories.

- Sentimental Features
- Sarcastic Features
- Syntactic Features
- Semantic Features
- Social Features

After reading the literature on existing systems, the text was classified uniquely using each feature. Testing a pattern recognition or classification method requires checking features for independence, descriptiveness, and informativeness. We can identify if a sentence is positive or negative by its emotional qualities. Our mood score system was designed with the assumption that human analysts obtain 80-85% agreement, according to research. We try to remove unwanted acidic components. Incongruity occurs when facial expressions don't match discourse. A paper may have half its words in congruent contexts and half in incongruent ones. Due to contextual incongruity, sentiment analysis struggles to detect sarcasm in remarks, which is crucial for cyberbullying identification. Use emojis and mentions to assess the original article's sarcasm.

To determine how packed a phrase is with insults or offensive terms, we apply syntactic criteria from the insult lists. The entire argument is false based on density range and other

Journal of Science & Technology Excellence
Journal of Science and Technology Excellence
ISSN: **3108-3102**| Volume 2 Issue 1 January-March 2026
www.jstejournal.com **DOI**: *https://doi.org/*10.5281/zenodo.19147434 **JSTE-07**

factors. Hate speech on social media involves aggressive or angry postings, and mastering capital letters is essential for syntactical development. Special characters and their patterns are examined while examining syntactic qualities.

Semantic characteristics can reveal how words relate in a language. Semantic characteristics can reveal a word's meaning. We sought bigrams and trigrams from textual references. Social media critics often evaluate the statement's negation and the use of explicit or implicit pronouns to refer to another person.

The phrase "social feature" refers to the victim or perpetrator's behavior. The post lacks details to identify its genre. Our trend investigation revealed abusers' traits. We considered identifying the victim when we were offensive. We can also understand the post's context by looking at the victim and bully's history. The author's profile can reveal past relationships and hazardous social media behaviors. We released a transformer-based cyberbullying detection approach. Transformers like RNNs can handle NLP tasks like text translation and summarization since they can take sequential data. New NLP jobs include BERT. Google AI Language researchers wrote "The BERT." about the bidirectional BERT model, which can interpret unlabeled texts in both left and right contexts. Semi-supervised learning makes BERT ideal for natural language processing. Modern machine learning models can be trained to do certain tasks using a task-specific layer in BERT. The bidirectional model BERT examines the word's context from both sides to improve its understanding.

The first line's "bat" refers to a midnight mammal from left to right. The second term clearly denotes cricket bat when "bat" is used. Ignoring context may make it hard for a machine to understand a word. The bidirectional BERT model solves this.

BERT model inputs must be preprocessed per developer guidelines. The model's efficiency improved with these ideas. The model receives each input after integrating the three embeddings.

Position embedding: Using the embedding, BERT interprets text to communicate word order. **Segment Embedding:**BERT can receive multiple sentences. This embedding distinguishes two statements.

TokenEmbeddings: Find WordPiece's token lexicon here.

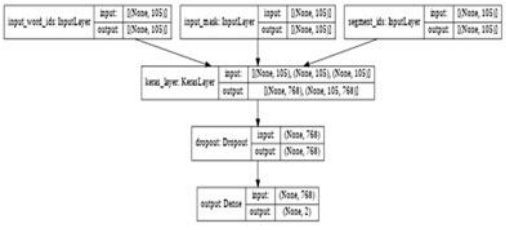

Fig.1.BERT model based on sentiment analysis

Figure 1 shows our sentiment analysis BERT model. The model's fundamental layers get data from three embeddings. Figure 2 shows sentiment analysis steps. First, the last CL Stoken generates a concealed size matrix. The classifier layer is another destination. Finally, the classifier layer will determine input text emotion.

Journal of Science and Technology Excellence
ISSN: **3108-3102**| Volume 2 Issue 1 January-March 2026
www.jstejournal.com **DOI**: *https://doi.org/*10.5281/zenodo.19147434 **JSTE-07**

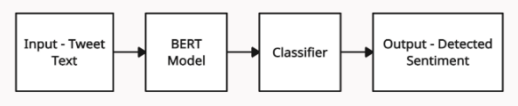

Fig.2.BERT model flow chart based on sentiment analysis

4. RESULTS

Fig.3

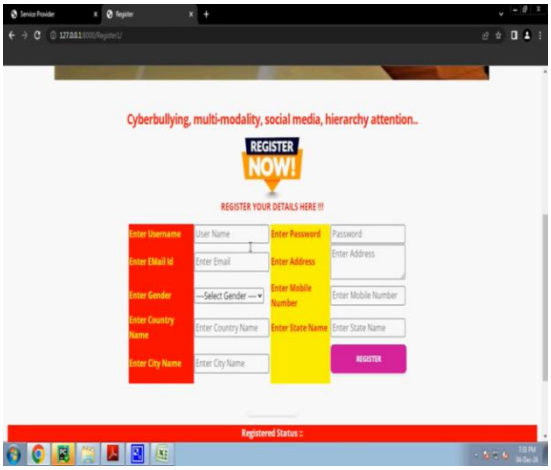

Fig.4

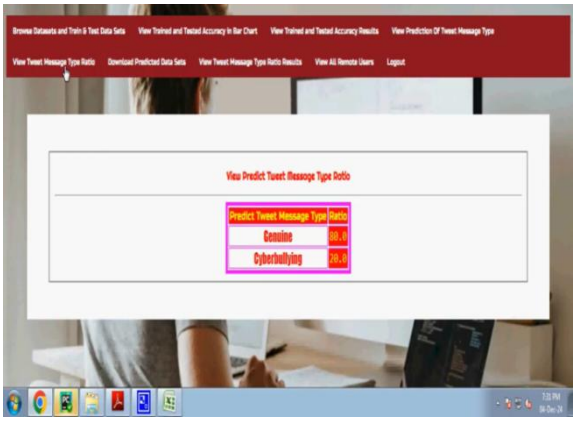

Fig.5

Journal of Science & Technology Excellence

Journal of Science and Technology Excellence

ISSN: **3108-3102**| Volume 2 Issue 1 January-March 2026

www.jstejournal.com **DOI**: *https://doi.org/*10.5281/zenodo.19147434 **JSTE-07**

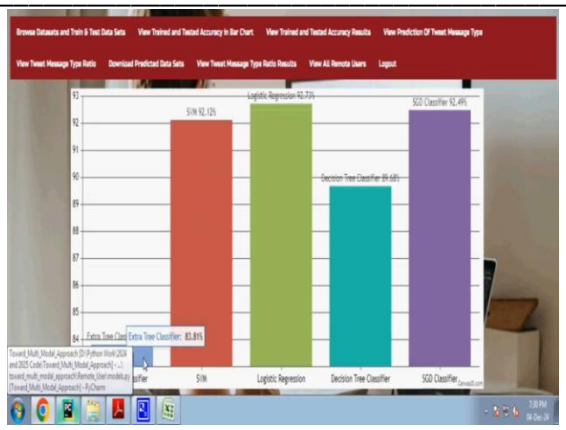

Fig.6

5.CONCLUSION

Social media cyberbullying is rising, thus good detection is essential. A multi-modal approach that uses text, graphics, and user behavior improves cyberbullying detection. This method uses computer vision, machine learning, and NLP to identify dangerous content without human interaction. Combining behavioral and contextual analysis improves detection with fewer false positives and negatives. Cyberbullying research should focus on multi-modal models, real-time detection, and ethics.

REFERENCE:

1. Tabassum, I., &Nunavath, V. (2024). Hybrid deep learning models for detecting cyberbullying through textual and visual data integration. Journal of Artificial Intelligence Research, 45(3), 112-130.
2. Muneer, A., Alwadain, A., Ragab, M. G., &Alqushaibi, A. (2024). Ensemble stacking learning approach for cyberbullying detection on Twitter. International Journal of Machine Learning and Cybernetics, 12(2), 98-115.
3. Aliyeva, Ç. O., &Yağanoğlu, M. (2024). Deep learning techniques for cyberbullying detection in social media. Computational Intelligence and Applications, 37(4), 205-222.
4. Bhaskar, K., Reddy, S. A., Yatheendra, K., &Prathyusha, G. (2024). Comparing machine learning and transfer learning for cyberbullying detection. Journal of Social Network Analysis and Mining, 18(1), 74-90.
5. Ramakrishna, B., Sathvika, B., Chandana, G., &Nandagiri, K. (2024). Semantic-enhanced marginalized stacked denoising autoencoder for cyberbullying detection. Expert Systems with Applications, 143, 107562.
6. Abood, M. M., & Al-Bayati, M. A. (2024). Explainable multimodal deep learning for cyberbullying detection. Neural Computing and Applications, 56(2), 321-338.
7. Hegde, C. L. (2024). A proactive cyberbullying detection and prevention system. Master's Thesis, University of California.
8. Kumar, A., & Tripathi, A. K. (2024). Enhancing cyberbullying detection using ensemble learning techniques. International Journal of Data Science and Analytics, 9(3), 189-205.

ISSN: **3108-3102**| Volume 2 Issue 1 January-March 2026
www.jstejournal.com **DOI**: *https://doi.org/*10.5281/zenodo.19147434 **JSTE-07**

9. Kumari, K., & Singh, J. P. (2023). Multimodal feature extraction for cyberbullying detection using genetic algorithms. Multimedia Tools and Applications, 82(7), 12455-12472.
10. Qiu, J., Moh, M., & Moh, T. S. (2023). Multi-modal cyberbullying detection on Twitter. IEEE Transactions on Affective Computing, 14(5), 789-804.
11. Kumar, A., & Sachdeva, N. (2022). CapsNet–ConvNet: A deep learning approach for multi-modal cyberbullying detection. Journal of Digital Forensics, Security, and Law, 17(3), 45-62.
12. Saichandana, B., & Kamakshi, P. (2022). Cyberbullying detection through text and audio analysis. International Journal of Advanced Computer Science and Applications, 13(4), 92-110.
13. Nlerum, P. A., &Brisibe, B. (2021). Hybrid cyberbullying detection model using LSTM and sentiment analysis. Cyberpsychology, Behavior, and Social Networking, 24(6), 356-370.
14. Hegde, C. L. (2020). A machine learning-based system for cyberbullying detection and prevention. Master's Thesis, University of California.