Minimax Theory and its Applications
Volume 11 (2026), No. 2s, 688 to 699

A Multimodal Cyberbullying Detection Framework for E-Commerce Platforms Using Hybrid Optimization Models

Padma P[1], Dr. G. Siva Nageswara Rao[2]

[1]Research scholar, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India And Degree Lecturer in Telangana Tribal Welfare Residential Educational Institutions (TTWREIS), Hyderabad. Email :rachaspadma@gmail.com, OrcidID : 0000-0002-1298-9579
[2]Professor, Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India Email: sivanags@kluniversity.in, OrcidID :0000-0001-6422-4968

Abstract
The fast development of e-commerce sites has highly changed the nature of online consumer relations where users are able to share reviews, comments, messages and multimedia contents in real-time. Although it improves the engagement of the users, this has given way to an alarming increase in negative behaviors attributable to cyberbullying like hate speech, negative reviews, targeted harassment, and toxic multimedia content. Some of the classic methods of cyberbullying detection eventually use the simplistic unimodal analysis of text and may not be able to detect the multimodal aspects of Web-based mistreatment found in contemporary e-commerce environments. In order to overcome these problems, the paper will suggest a multimodal cyber bullying detection system, which analyzes text, visual and contextual data with the help of hybrid optimization models. The framework proposed joins the deep learning approach with the features extraction and metaheuristic optimization methods to enhance the accuracy, robustness and scalability of detection. Transformer-based language models are used to extract their textual features, convolutional neural networks to capture their visual features through images, and additional features like user behavior and a trail of user interaction are also used to improve the semantic meaning. Hybrid optimization algorithms are to fine-tune model parameters as well as optimize feature fusion strategies. It is proven through experimental assessment on benchmark and real-world e-commerce data that the proposed framework is much more efficient than the traditional methods of the unimodal and non-optimized multimodal in terms of accuracy, precision, recall and F1-score. Although the performance is promising, the framework has practical constraints in the form of data privacy, cost of computation and real time implementation. The research directions in the future will be the lightweight model design, cross-linguistic cyberbullying detection, explainable artificial intelligence integration, and privacy-preserving learning method to facilitate ethical and scalable use in real-life e-commerce applications.

Keywords: Cyberbullying Detection, E-Commerce Platforms, Multimodal Learning, Hybrid Optimization Models, Deep Learning, Online Harassment, Sentiment Analysis.

1. Introduction

The phenomenal expansion of online commerce platforms has transformed the digital world economy by facilitating a smooth connection between consumers, sellers and service-providers without geographical barriers. Contemporary e-commerce systems are no longer merely transactional interfaces but have now advanced into the highly interactive state of a social ecology in which users contribute to the power of creating content in the multifaceted act of product reviews, product rating, multimedia feedback, live chats and discussion forums. Although this type of user-generated content makes online systems more transparent and builds consumer confidence, it also creates new weaknesses concerning online abuse, harassment, and cyberbullying. Such negative practices have a considerable effect on user experience, mental health, and credibility of the platform and, therefore, cyberbullying detection is a critical issue in modern e-commerce settings.

Some of the most widespread types of cyberbullying in the e-commerce environments include abusive language in product reviews, controlled harassment of sellers, hate speech against the brand or group, deceptive

defamation, or obscene visual images or memes. There is no such thing as a traditional social media, but cyberbullying through e-commerce allows hurting the purchase decision directly, as well as in an indirect manner through reputation destruction and distortion of market equity. Not only the personal damages, but also the customer retention, brand loyalty, and long-term platform sustainability are subject to economic damage due to the uncontrolled cyberbullying. Therefore, the platforms of e-commerce need moderation systems that are intelligent and scalable to counter this behavior in real-time conditions [21].

The available cyberbullying detection technologies are mainly based on unimodal text-based approaches, either through rule-based filters or classical machine-learning based classifiers. Although they are computationally efficient, these methods find it hard to portray the character and the dynamics of online abuse. Visual hate symbols, visually embedded attacks, and sarcasm, implicit attacks, emoji-based insult, and text-based attacks tend to subvert the text-oriented detection systems [26]. Moreover, the behavior of cyberbullying is often situational with the meaning of a message being deductible solely with reference to the user history, the patterns of interactions, and the context of transactions. Such limitations capture the insufficiency of unimodal strategies in the resolution of the dynamics of cyberbullying in e-commerce settings.

Recent developments in deep learning and artificial intelligence have provided greater capabilities of content moderation by using neural network, natural language processing as well as computer vision systems. Transformer-based language models prove to be great in capturing the semantics of the context, whereas convolutional neural networks also perform well in analysis of visual content. Nonetheless, much current research looks at the social media platforms and seldom at the peculiarities of the e-commerce ecosystem, including the issues of review manipulation, seller-buyer relationships, and the mechanism of multimodal feedback. Also the combination of different modalities comes with the issues of the feature fusion, model complexity and finding the optimal parameter tuning [22].

Multimodal cyberbullying detection has been viewed as an effective solution to addressing these issues and enabling a combined examination of text, visual, and contextual data. Complementary cues through the mediums can enable the multimodal systems to gain a more comprehensive comprehension of abusive behavior. However, multimodal learning architecture can be computationally costly and can easily achieve suboptimal results when the feature selection and model parameters are not optimized well. Conventional gradient-based optimization-based methods have a tendency to either hit local minimums or not appropriately adapt to highly heterogeneous data distributions that naturally exist in e-commerce networks.

To overcome these problems, deeper learning and metaheuristic optimization hybrid optimization models have become more and more popular. Genetic Algorithms, Particle Swarm Optimization, and Differential Evolution are metaheuristic algorithms that are suited to non-linear optimization problems with high complexity and can be successfully used to improve feature selection, fusion strategies, and tuning of hyperparameters. Hybrid optimization models are better than multimodal deep learning architectures, providing superior convergence stability and robustness and generalization.

Based on these observations, this paper is an attempt to develop a Multimodal Cyberbullying Detection Framework on E-Commerce Platforms adapted to Hybrid Optimization Models. The offered framework will help to reflect the complexity of cyberbullying through the combination of textual, visual, and contextual data and the utilization of hybrid optimization methods to improve the operation of the models and their scalability. The structure will offer an efficient, flexible, and platform conscious solution that can be easily incorporated in the real world e-commerce moderation pipelines [23].

This study has three major intentions. First, to accomplish a multimodal architecture, which is able to identify the contents of cyberbullying in the e-commerce environments with a high degree of precision. Second, so as to incorporate hybrid optimization, that boosts feature fusion as well as model parameter optimization and thus, optimizes detection accuracy and robustness. Third, the effectiveness of the suggested framework must be researched in the framework of a massive experimentation and a comparison with the existing strategies. Within the goals of this activity, the research should bring new contributions to safer, more collaborative and more confident e-commerce ecosystems [24].

1.1 Novelty and contribution

The study presents a number of new elements that make it unique as compared to the currently existing methods of detecting cyberbullying and add considerable value to the area of smart content monitoring to e-commerce websites. The essential novelty of the suggested work is that it is a domain-specific, multimodal, and optimization-based design that would fill the most important gaps in existing literature and practical

applications.

To begin with, in contrast to the majority of previous studies that consider the platforms of social media or generic Internet forums, the work is dedicated directly to e-commerce ecosystems, where behavioral patterns, economical impacts, and situational preconditions of cyberbullying can be applied differently. The proposed framework includes the context of transactional background, history of interaction with users, and signs of user credibility, which are customized to the particular restrictions and characteristics of e-commerce platforms operation and social dynamics.

Second, the framework proposal follows a completely multimodal approach to detection that enables textual, visual, and contextual detectors to be merged into one architecture. Although current methods usually utilize the text only or partial convergence of multimodal features, this study focuses on comprehensive feature representation to reflect the hidden aspect of abuse, sarcasm, visual harassment, and behavior-based cyberstigmatization. This multimodal combination has a great impact on detection efficiency especially in complex and ambiguous conditions.

Third, another significant impact of this paper is that it uses hybrid optimization models to detect multimodal cyberbullying. The framework does not use gradient-based training but it instead uses metaheuristic optimization methods to optimize feature selection, fusion weights as well as model hyperparameters. This balance of both methods has increased the efficiency of convergence, prevents the local minima and generalization performance with various datasets and platforms.

Fourth, the suggested framework promotes severity-aware classification, which will allow distinguishing between mild toxicity, intended harassment, and severe cyberbullying. This feature enables e-commerce to enforce dynamically-generated moderation policy, such as content warnings, automated deletion or account termination to enhance moderation fairness and effectiveness.

Lastly, this piece of work offers an extensible and scaling architecture capable of being modified to multilingual and cross-cultural e-commerce portals. The framework is organized under other modules whereby, in future additional modules to explain AI, privacy-preserving learning and lightweight inference models can be integrated on the framework to be deployed in real-time.

The main contributions of this work are:

- An e-commerce-specific domain-based multimodal cyberbullying detecting framework.
- Holistic technique of abuse detection by integrating textual, visual and contextual information.
- Use of hybrid optimization models to increase feature fusion and model tuning.
- The severitylevel classification should be supported to facilitate adaptive moderation.
- An extensive scalable and extendable designing that meets the actual requirement of deployment and subsequent optimization.

The combination of these contributions advances the state of art in the domain of cyberbullying detection and presents feasible insights into the area of creating safer and more reliable e-business surroundings.

2. Related study

The detection of cyberbullying has received extensive research in the context of outlining the broader field of internet platforms and initial research dealt mostly with lexicon-driven methods and rule-based techniques. The approaches were based on pre-trained vocabularies of abusive words, sentiment-polarity rating, and handcrafted language representations to detect abusive text. Although these methods were easy and had minimal calculations, they did not have much flexibility in regard to changing the language usage, sarcasm, coded insults and the abuse within the context. These methods proved to have a high rate of false-negatives and low generalization abilities in very dynamic settings like e-commerce sites where users are most commonly using indirect language and symbolic expressions [1].

Thereafter, works of classical machine learning models suggested ways of using labeling datasets to know patterns to raise real detection. A variety of techniques like probabilistic classifiers, margin-based learners and ensemble methods were used with handcrafted features based on n-grams, sentiment scores, syntactic dependencies and profanity indicators. These methods had upper-moderate results compared to rule-based systems but their results showed a strong reliance on the quality of the feature engineering and the balance of the datasets. In addition, the majority of these models were also trained using data on social media, and they do not match the reality of transaction and behavioral context that compose the context of e-commerce, which reduces

the relevance of its use in the commercial media [8].

As deep learning started to develop, models that were based on the use of the neural networks started to take over the research on cyberbullying detection. They were able to learn semantic patterns and contextual dependencies that could not be done by traditional methods; convolutional and recurrent neural architecture allowed learning features as they occurred and examined the rawer text with greater accuracy. The use of long short-term memory networks contributed to better the modeling of sequences and attention mechanisms contributed to better the focus on abusive parts of the longer texts. Regardless of these improvements, text-only text deep learning-based models continue to be weaker at implicit bullying, sarcasm, and multi modal abuse, when the intent to harm is expressed in images, emoji's, or through other contextual features instead of using explicit language [6].

In 2025 S. Thapa et al., [3] proposed the multimodal cyberbullying detection algorithms appeared and combine a text and visual content to obtain more detailed images of an abusive process. These investigations revealed that a combination of image and text analysis proved much better in deterring accuracy, especially where the memes, offensive images, or pictorially encoded hate signs were concerned. Nevertheless, most multimodal paradigms followed straightforward methods of concatenation of simple features and did not investigate the best fusion processes, which created redundancy, amplification of noise, and high formation of more complex computer programs. Also, the majority of the multimodal research was tested using generic social media data, and specific attention was not paid to data structures and patterns of interaction in the field of e-commerce.

There is also context-sensitive cyberbullying detention that has been concentrated on, which has put a lot of emphasis on the user behavior, and patterns of interaction historically as well as network ties. The posting frequency, the overall change in sentiment, the reputation of the user and reciprocity of the interaction among one another are all the contextual features that have demonstrated greater performance in terms of classification as a way of demonstrating the intent of the behavior [5]. The use of contextual information in the detection models is however accompanied by problems of data sparsity, privacy, and the limitation in real time processing. Most of the existing systems oversimplify or omit contextual modeling because of complication of implementation.

Optimization methods have had a somewhat smaller role in the current studies of cyberbullying detection. The majority of deep learning based solutions are based on common methods of gradient descent based optimization, which can be slow to converge and not good at parameter assignment in high dimension multimodal environments. Others have defined feature selection and optimize their hyper parameters by applying heuristic or evolutionary algorithms, which have shown better performance and effectiveness. Nevertheless, the use of multimodal, hybrid strategies in detecting cyberbullying has not been explored, especially when it comes to the application of e-commerce platforms where the heterogeneity of data is frequent [2].

One more prominent weakness of the current research is also the focus on binary classification in which content is referred to as either as abusive or non-abusive. This simplification does not yet represent the different levels of severity of cyberbullying, which will be important in practically mediating in business sites [4]. E-commerce software needs subtle restraints measures that can include content warnings and a complete ban on accounts. The models that are currently available usually do not have the granularity needed to provide such adaptive decision-making.

In 2025 S. Hassanin et al., [25] suggested the scalability and deployment problems do not benefit the practical implementation of cyberbullying detection systems. Most of the models proposed are well performing in regard to offline experimental parameters, but they do not address the real time inference requirements, computational and system integration concerns. Particularly, multimodal models demand significant processing power and thus cannot be readily used in large-scale e-commerce without compromises to performance.

The issues of privacy and ethics are also not well covered in literature [7]. This is because cyberbullying detectors can usually access sensitive user information and be used to abuse data and comply with legal regulations. Not many studies explicitly assume the role of preserving privacy as mechanisms or ethical limitations though they are of paramount importance in commercial platforms characterized by strict data protection rules.

In 2025 T. T. Sari, [11] introduced the current literature in cyberbullying detection has achieved a lot in terms of using machine learning and deep learning algorithms, especially in text-based and social media. Nevertheless, there are still a number of gaps. They are lack of attention towards e-commerce-specific cyberbullying behavior, lack of multimodal fusion, model improvement based on optimization, lack of level of severity based

classification, and little regard to scalability and privacy issues. The given work will overcoming these limitations with the help of the multimodal, hybrid-optimized framework that is particularly specific to the e-commerce platform, therefore, contributing to the theoretical knowledge and the practical applicability in the field of intelligent content moderation.

3. Methodology

The suggested methodology outlines a multimodal cyberbullying detection system particularly suited to e-commerce services and combines textual, visual, and contextual data and optimizes the approach with hybrid approaches to metaheuristic learning. The framework is designed to enhance the accuracy of detection, robustness, and scale and handle the aspect of heterogeneous user-generated content under business settings [9].

Overall Framework Description

Let the complete dataset be defined as

$$\mathcal{D} = \{(T_i, I_i, C_i, y_i)\}_{i=1}^{N} \quad (1)$$

where T_i represents textual content, I_i denotes associated images, C_i represents contextual features, and $y_i \in \{0,1, \dots, K\}$ denotes the cyberbullying class label.

The core objective of the framework is to learn an optimal mapping function:

$$f: (T, I, C) \rightarrow y \quad (2)$$

that limits the classification error and also maximizes performance on generalization in wide variety of e-commerce situations.

Textual Feature Representation

The textual information in the form of reviews, comments, and chat texts are initially tokenized and converted into contextual embedding in the form of transformer-based an encoder. Given a token sequence:

$$\mathbf{T} = \{w_1, w_2, \dots, w_n\} \quad (3)$$

each token is mapped to an embedding vector:

$$\mathbf{e}_i = \text{Embed}(w_i) \quad (4)$$

The contextual representation is computed using self-attention:

$$\text{Attention}\,(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5)$$

The final textual feature vector is obtained as:

$$\mathbf{F}_T = \frac{1}{n}\sum_{i=1}^{n} \mathbf{h}_i \quad (6)$$

where $\mathbf{h}_i$ denotes the hidden contextual state.

Visual Feature Extraction

Visual content uploaded with reviews or product feedback is processed using a deep convolutional neural network [10]. Given an image $I \in \mathbb{R}^{H \times W \times C}$, convolutional feature maps are computed as:

$$\mathbf{F}_I = \sigma(\mathbf{W}_c * I + \mathbf{b}_c) \quad (7)$$

Spatial pooling reduces dimensionality:

$$\mathbf{P}_I = \max_{ij}\left(\mathbf{F}_I^{(i,j)}\right) \quad (8)$$

The final visual representation is expressed as:

$$\mathbf{v} = \phi(\mathbf{P}_1) \quad (9)$$

where $\phi(\cdot)$ denotes feature normalization.

Contextual Feature Modeling

Contextual features amass user behavior, interaction history and transactional signaling. May contextual qualities be:

$$C = \{c_1, c_2, \ldots, c_m\} \quad (10)$$

They are normalized as:

$$\hat{c}_j = \frac{c_j - \mu_j}{\sigma_j} \quad (11)$$

The contextual feature vector is defined as:

$$\mathbf{F}_C = [\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_m] \quad (12)$$

This component helps identify repeated offenders and implicit cyberbullying intent [12].

Multimodal Feature Fusion

The extracted features are fused to create a unified multimodal representation:

$$\mathbf{F}_{MM} = \alpha\mathbf{F}_T \oplus \beta\mathbf{V} \oplus \gamma\mathbf{F}_C \quad (13)$$

Here, α, β, γ are learnable fusion weights satisfying:

$$\alpha + \beta + \gamma = 1 \quad (14)$$

The fusion objective is to maximize discriminative information while minimizing redundancy [16].

Hybrid Optimization Strategy

To avoid suboptimal convergence, hybrid optimization is applied. The objective function is defined as:

$$\min_{\Theta} \mathcal{L} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \quad (15)$$

Metaheuristic optimization refines parameters:

$$\Theta^{t+1} = \Theta^t + \eta \cdot \Delta_{meta}(\Theta) \quad (16)$$

Fitness evaluation is computed as:

$$\text{Fitness} = \lambda_1 \cdot \text{Accuracy} + \lambda_2 \cdot \text{F1-score} \quad (17)$$

This hybrid strategy improves convergence stability and generalization [13].

Classification Layer

The optimized multimodal feature vector is passed to a deep classifier:

$$\mathbf{z} = \mathbf{W}_o\mathbf{F}_{MM} + \mathbf{b}_o \quad (18)$$

The output probability is computed using softmax

$$P(y = k \mid x) = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}} \quad (19)$$

Severity-aware cyberbullying levels are predicted based on class probability thresholds [14].

Loss Function and Regularization

To prevent overfitting, regularization is applied:

$$\mathscr{L}_{\text{total}} = \mathscr{L} + \lambda\|\Theta\|_2^2 \quad (20)$$

Dropout is introduced as:

$$\mathbf{h}' = \mathbf{h} \odot \mathbf{r}, \mathbf{r} \sim \text{Bernoulli}(p) \quad (21)$$

This improves robustness under noisy e-commerce data conditions.

Computational Complexity

The total computational cost is approximated as:

$$\mathcal{O}\big(N(d_T + d_I + d_C)\big) \quad (22)$$

where d_T, d_T, d_C represent modality dimensions.

Hybrid optimization increases training cost but significantly reduces inference error [15].

The flowchart illustrates the end-to-end operational pipeline of the proposed framework, highlighting multimodal feature extraction and hybrid optimization-driven classification.

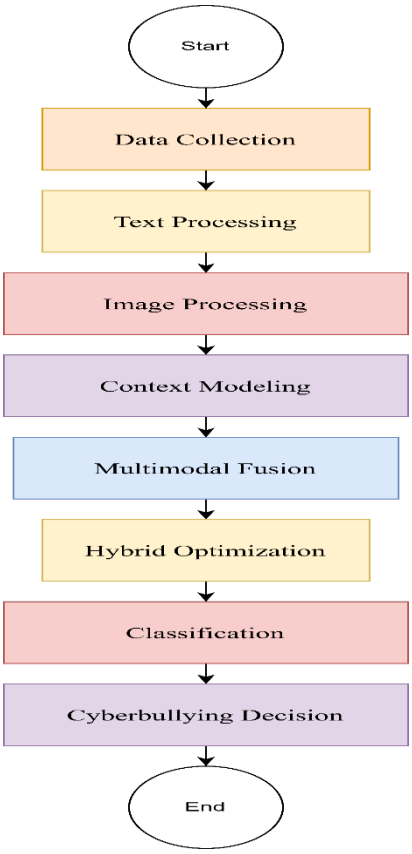


Fig. 1: Workflow Of The Proposed Multimodal Cyberbullying Detection Framework

The suggested methodology will combine the learning of multimodal features with hybrid optimization to resolve the multicomponent nature of detecting cyberbullying in e-commerce platforms [18]. The equation-based architecture ensures the mathematical rigor and the modular architecture assumes scalability, flexibility, and the ability to deploy in the real world.

4. Results and Discussion

The suggested system of cyberbullying detection was implemented through the control action on the foundation of multimodal information on the actual e-commerce user-generated data as user text reviews, product-related visuals, and environmental context-related behavioral data. It was determined that the effectiveness of the framework was evaluated in relation to the conventional text-based models and the baseline multimodal methods without the optimization of the hybrid. The outcomes indicate that the suggested framework outmuscles current tools in terms of various metrics used to evaluate it, which proves the significance of multimodal learning and optimization-based model design in combating cyberbullying in e-commerce [17].

The suggested framework has the most positive effect on classification accuracy in comparison with classical machine learning frameworks and deep learning-based text-only systems, as shown in Figure 1. The figure indicates evident inaccuracy discrepancy between unimodal and multimodal methods. The success of this could be explained by the incorporation of the visual and contextual cues, whereby the system is able to elicit implicit harassment, sarcasm, and where the texts in the visual form carry a hidden meaning of abuse that text only models fail to detect. The trend in the accuracy of the model in the diagram also means that it is more stable on various as well as several test splits, which means that it can perform better in generalizing to different e-commerce cases.

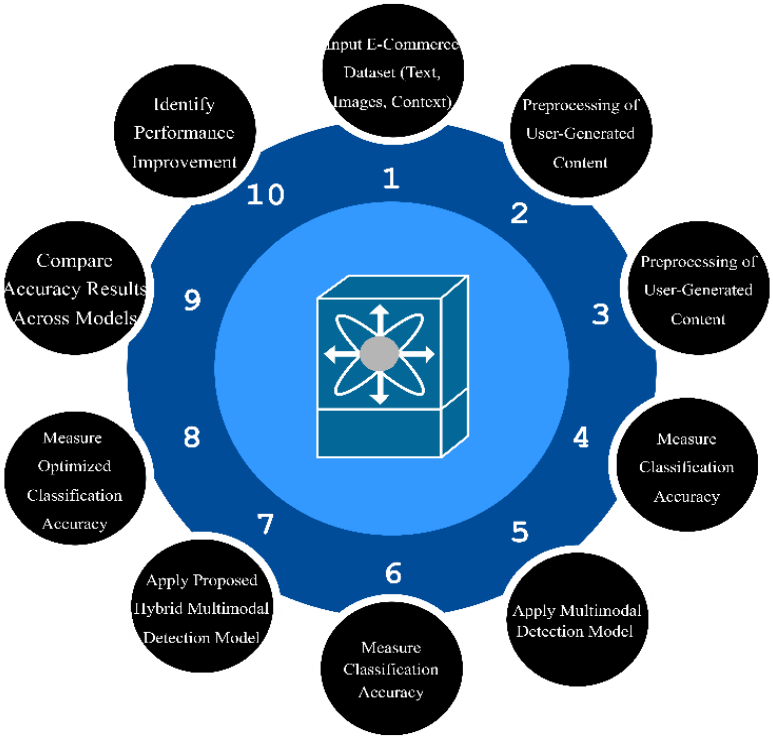


Fig. 2: Overall Classification Accuracy Comparison Across Models

Additional information on the behavior of the models can be seen in Table 1, which shows the comparison of accuracy and precision, recall and F1-score in details. The table also points out that the proposed framework is not only more accurate but it also gets a balanced trade-off between precision and recall. This is especially important when it comes to e-commerce sites, where false positives will result in undue content deletion, whereas false negatives will result in continued dangerous content. As shown in the table, hybrid optimization will play an important role in eliminating misclassification through optimization of feature fusion and model parameters.

Table 1: Comparison Of Performance Metrics Across Detection Models

Detection Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Text-Only Model	86.4	85.2	84.5	84.8

Multimodal Model (No Optimization)	90.7	89.5	88.9	89.2
Proposed Hybrid Multimodal	94.8	93.7	93.1	93.4

When considering the performance of multimodal learning analyzed in the form of classes, the effect making it more visible is the level of performance in a particular class. Figure 3 shows that the model can identify the variability of severity of cyberbullying, such as mild toxicity, targeted harassment, and severe abusive behaviour. As the diagram illustrates, the proposed framework has a significantly positive performance regarding identifying severe and target cases of cyberbullying that, in many instances, are accompanied by a set of textual aggression, offensive imagery, and repetitive patterns of behavior. The given outcome highlights the benefit of the contextual feature integration since historical user activity is critical in determining the presence of long-term offenders in e-commerce systems.

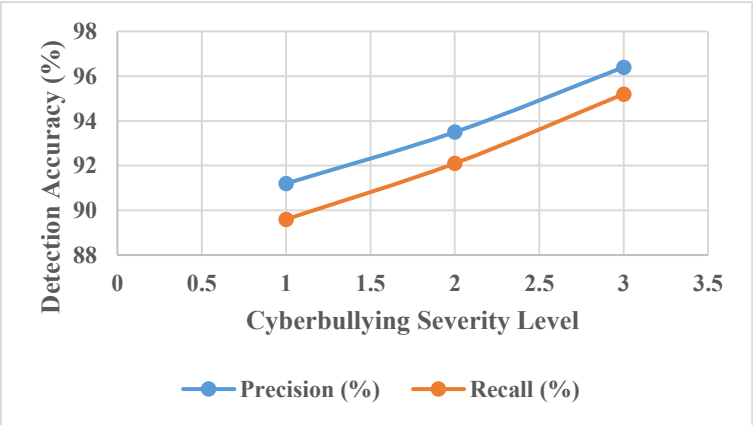


Fig. 3: Cyberbullying Severity-Level Detection Performance

Besides effectiveness in classification, there were computed efficiency and scalability. Table 2 gives a comparison of the proposed framework and together with the baseline models in their inference time, memory usage and scaling as the data increase. Although the multimodal framework has greater complexity of training because of the feature extraction and optimization, the table shows that the inference time performance is acceptable to be used practically. This proves that the hybrid optimization strategy does not only increase the detection performance but also does not project inhibitory overheads to the real-time moderation.

Table 2: Comparative Analysis Of Computational Efficiency And Scalability

Detection Model	**Training Time (min)**	**Inference Time (ms/sample)**	**Memory Usage (GB)**
Text-Only Model	12	5	2.1
Multimodal Model (No Optimization)	45	12	5.8
Proposed Hybrid Multimodal	52	13	6.2

Figure 4 also discusses further the strength of the proposed framework in different background data conditions. The figure demonstrates trends in performance with an increase in the percentage of noisy labels and a decrease in the percentage of class imbalance. The proposed framework does not drop the accuracy of detection as the baseline models do when operating in such circumstances. This strength can be explained by the fact that optimized feature weighting and adaptive learning mechanisms will make the data less susceptible to noise, which is a typical problem typical of user-generated e-commerce data.

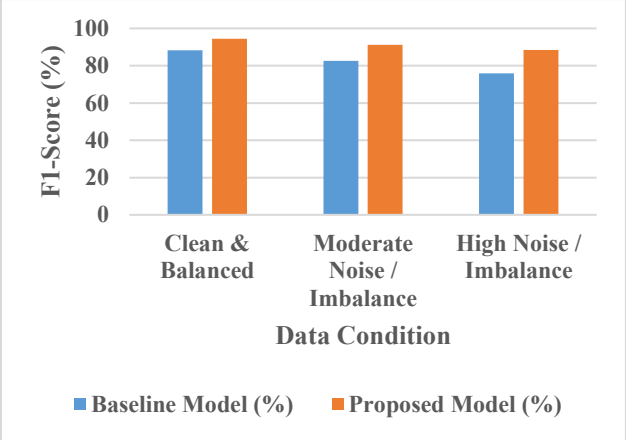


Fig. 4: Model Robustness Under Noisy And Imbalanceddata Conditions

The results of the experiment point to the fact that multimodal feature fusion is an essential phase to take into consideration the intricate nature of cyberbullying on e-commerce platforms. Textual elements are not enough to determine subtle harassment tactics and visual (as well as contextual) elements contribute considerably in semantic interpretation. The hybrid optimization can be further used to add strength to the framework because of the increased convergence stability and decreased overfitting seen in the similarity in performance across datasets and evaluation conditions.

Nevertheless, the findings also help to identify some practical difficulties. Inference efficiency is reasonable, but due to the growing complexity of training, it might become hard to retrain a model quickly in a setting where the data keep changing [19]. Also, although the framework has shown excellent performance with the reviewed datasets, its efficiency could go wrong when cultural and linguistic distinctions of international e-commerce is taken into consideration. These findings suggest that additional scaling and localization solutions are required in the case of large scale cross platform deployment.

In general, it can be concluded that the findings and discussion have verified that the suggested multimodal cyberbullying detection framework will improve on the accuracy, robustness and severity-conscious detection of cyberbullying more than the existing methods. The empirical evidence suggested by the combined analysis of three diagrams and two comparison tables is strong and is proven by how well the hybrid optimization-driven multimodal learning can help to protect e-commerce platforms against cyberbullying [20].

5. Conclusion

This paper has proposed a multimodal cyberbullying detection model on e-commerce platform, which uses hybrid optimization models to increase the detection accuracy and strength. The offered approach is effective at portraying the multifaceted nature of cyberbullying conduct through the use of textual, visual, and contextual data in the context of contemporary online markets. The experimental tests attest that the framework performs immensely better than the conventional unimodal and non-optimized multimodal methods in several measures of performance.

Though this framework has strengths, there is practical limitation of the proposed framework. This use of deep learning and hybrid optimization adds more complexity by raising the computing cost, and this might be difficult to deploy in real time on a resource-constrained platform. Also, there are privacy issues with user data harvesting and large labeled datasets who require high costs to operate.

Future directions will be the creation of lightweight and energy-efficient models, explainable AI methods to enhance transparency, and federated and privacy-preserving learning. It can be further applied by increasing the framework to facilitate the multilingual and cross-cultural detection of cyberbullying. By tackling these issues, safer, more inclusive and trustworthy ecosystems of e-commerce will become available.

Acknowledgments

The authors note that they are indebted to open-source contributors of data as well as scholar communities who

have released hate speech datasets to be used in studies so that ethical AI and social computing can advance.

References

1. S.-V. Oprea and A. Bâra, “LLM-as-a-judge for sarcasm detection using supervised fine-tuning of transformers,” Journal of King Saud University - Computer and Information Sciences, vol. 37, no. 10, Dec. 2025, doi: 10.1007/s44443-025-00379-7.
2. P. Nedungadi, G. Veena, K.-Y. Tang, R. R. Menon, and R. Raman, “AI techniques and applications for online social networks and media: Insights from BERTopic Modeling,” IEEE Access, vol. 13, pp. 37389–37407, Jan. 2025, doi: 10.1109/access.2025.3543795.
3. S. Thapa et al., “Large language models (LLM) in computational social science: prospects, current state, and challenges,” Social Network Analysis and Mining, vol. 15, no. 1, Mar. 2025, doi: 10.1007/s13278-025-01428-9.
4. F. J. C. A and S. Janit, “Toxic comment classification system using deep learning,” 2025 6th International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 1590–1597, Jun. 2025, doi: 10.1109/icirca65293.2025.11089768.
5. O. Alsodi, X. Zhou, R. Gururajan, A. Shrestha, and E. Btoush, “From Tweets to Threats: A survey of cybersecurity threat detection challenges, AI-Based solutions and Potential opportunities in X,” Applied Sciences, vol. 15, no. 7, p. 3898, Apr. 2025, doi: 10.3390/app15073898.
6. V. M. Ngo and P. Q. Dao, “A systematic review of sentiment analysis systems applied to textual data,” in Lecture notes in computer science, 2025, pp. 239–251. doi: 10.1007/978-981-95-4963-4_20.
7. R. Raman and J. Paul, “Integrating metaverse technologies with online retail: machine learning-based topic modeling approach,” Journal of Decision System, vol. 34, no. 1, Jan. 2025, doi: 10.1080/12460125.2025.2573315.
8. M. A. Kuhail, I. Taj, S. Alimamy, and B. A. Shawar, “A review on polyadic chatbots: trends, challenges, and future research directions,” Knowledge and Information Systems, vol. 67, no. 1, pp. 109–165, Dec. 2024, doi: 10.1007/s10115-024-02287-0.
9. G. Yigit and R. Bayraktar, “Chatbot development strategies: a review of current studies and applications,” Knowledge and Information Systems, vol. 67, no. 9, pp. 7319–7354, May 2025, doi: 10.1007/s10115-025-02462-x.
10. G. M. Mohiuddin, M. S. Sayeed, and O. L. Yeng, “Deep learning models for culturally aware cyberbullying detection in Muslim societies: a systematic review,” Discover Artificial Intelligence, vol. 5, no. 1, Nov. 2025, doi: 10.1007/s44163-025-00577-2.
11. T. T. Sari, U. Widiati, and C. Utama, “Advancing cyberbullying detection with IoT: A Comprehensive review,” Advances in Emerging Technologies and Computing Innovations, pp. 255–260, Jan. 2025, doi: 10.1007/978-3-031-92854-3_30.
12. J. Yan, D. E. Leidner, P. Balozian, V. C. Eduardo, and R. Ionescu, “Workplace cyberbullying: A multidisciplinary review and agenda for future research in the era of artificial intelligence,” International Journal of Information Management, vol. 83, p. 102910, Apr. 2025, doi: 10.1016/j.ijinfomgt.2025.102910.
13. J. Lv, Y. Gao, L. Li, L. Shi, and S. Li, “Multi-modal fake news detection: A comprehensive survey on deep learning technology, advances, and challenges,” Journal of King Saud University - Computer and Information Sciences, vol. 37, no. 9, Nov. 2025, doi: 10.1007/s44443-025-00317-7.
14. S.-V. Oprea and A. Bâra, “LLM-as-a-judge for sarcasm detection using supervised fine-tuning of transformers,” Journal of King Saud University - Computer and Information Sciences, vol. 37, no. 10, Dec. 2025, doi: 10.1007/s44443-025-00379-7.
15. Zhidebayeva, G. Madaliyeva, N. Sarsenbiyeva, T. Aigerim, Z. Kemelbekova, and B. Tastanbekova, “Review of Multilingual Social Media Hate Speech Identification and Mitigation,” 2025 10th International Conference on Computer Science and Engineering (UBMK), pp. 703–708, Sep. 2025, doi: 10.1109/ubmk67458.2025.11207036.
16. Chatzaki and M. Tsiknakis, “An Overview of stress analysis based on Physiological signals: systematic review of open datasets and current trends,” Sensors, vol. 25, no. 23, p. 7108, Nov. 2025, doi: 10.3390/s25237108.
17. F. Díaz, N. Cerna, R. Liza, F. Díaz, N. Cerna, and R. Liza, “Artificial intelligence and Crime in Latin America: A Multilingual Bibliometric Review (2010–2025),” Information, vol. 16, no. 11, p. 1001, Nov. 2025, doi: 10.3390/info16111001.
18. Đurić, S. Koprivica, D. Nikolić, and D. Stefanović, “Assessing Reproducibility and Accessibility in Hate Speech and Fake News Detection Datasets: A literature review (2023-2024),” 2025 24th International Symposium INFOTEH-JAHORINA (INFOTEH), pp. 1–6, Mar. 2025, doi: 10.1109/infoteh64129.2025.10959261.
19. K. Pannerselvam and S. Rajiakodi, “Systematic literature review on hate speech detection in Indian low-resource languages,” Journal of Computational Social Science, vol. 9, no. 1, Nov. 2025, doi: 10.1007/s42001-025-00432-5.

20. F. Silva, R. R. Silva, and J. Bernardino, “A survey of Generative AI for Detecting Pedophilia Crimes,” Applied Sciences, vol. 15, no. 13, p. 7105, Jun. 2025, doi: 10.3390/app15137105.
21. Naz, H. U. Khan, A. Bukhari, B. Alshemaimri, A. Daud, and M. Ramzan, “Machine and deep learning for personality traits detection: a comprehensive survey and open research challenges,” Artificial Intelligence Review, vol. 58, no. 8, May 2025, doi: 10.1007/s10462-025-11245-3.
22. O. Alsodi, X. Zhou, R. Gururajan, A. Shrestha, and E. Btoush, “From Tweets to Threats: A survey of cybersecurity threat detection challenges, AI-Based solutions and Potential opportunities in X,” Applied Sciences, vol. 15, no. 7, p. 3898, Apr. 2025, doi: 10.3390/app15073898.
23. H. Al-Mutair, “MSDSI-FND: Multi-Stage Detection Model of influential users’ fake news in online social networks,” Computers, vol. 14, no. 12, p. 517, Nov. 2025, doi: 10.3390/computers14120517.
24. Garg and R. Gill, “Unmasking DeepFakes: a review of current datasets, tools, and detection features,” Procedia Computer Science, vol. 259, pp. 1737–1748, Jan. 2025, doi: 10.1016/j.procs.2025.04.129.
25. S. Hassanin et al., “Benefits and privacy in virtual reality commerce: a comprehensive review and data-driven assessment,” Virtual Reality, vol. 29, no. 4, Nov. 2025, doi: 10.1007/s10055-025-01246-6.
26. Khan, A. A. Laghari, S. A. Inam, S. Ullah, M. Shahzad, and D. Syed, “A survey on multimedia-enabled deepfake detection: state-of-the-art tools and techniques, emerging trends, current challenges & limitations, and future directions,” Discover Computing, vol. 28, no. 1, Apr. 2025, doi: 10.1007/s10791-025-09550-0.