International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

JUSTPING: Guarding Against Online Predators

Neha Beegam P.E[1], Aswathy Ravindran[2], Elna Thankachan[2], Gowri Saadhika[2], Mariya Nixon[2]

[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Vazhakulam, Kerala, India

[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Vazhakulam, Kerala, India

ARTICLEINFO

Article History:
Published : 05 April 2025

Publication Issue :
Volume 12, Issue 12
March-April-2025

Page Number :
228-240

ABSTRACT

With the increasing cybercrime on social media, this study creates a strong security system based on sophisticated machine learning and behavioral techniques. It trains on labeled image datasets to identify explicit content and text datasets to recognize cyberbullying, depressive language, and off-topic posts. The main features include identifying inappropriate images, iden- tifying cyberbullying, classifying user comments, and recognizing depressive language for mental health assistance. The system also guards against misuse of images by continuous notifications. Using deep learning architectures such as CNN, LSTM and NLP, and combining them within an Android-based application, this solution increases user safety and authenticity on social networks.

Index Terms—Natural Language Processing, Long Short- Term Memory, Convolutional Neural Networks.

INTRODUCTION

Social media has been growing rapidly, transforming how people connect and interact within society. However, this growth has also led to a rise in cyber crimes, including harassment, cyberbullying, and the distribution of explicit materials. These activities pose significant risks to users [1], [2]. Addressing these issues requires innovative strategies that can effectively mitigate risks while maintaining the benefits of technology-driven connections.

The project proposes an enhanced solution aimed at prevent- ing both existing and potential online threats. By continuously monitoring a data set of text generated by users on the social platform, the system identifies potentially harmful behaviors, such as sharing explicit content, engaging in cyberbullying, and using negative emotional language indicative of mental health issues such as depression [3], [4]. This approach rep-resents a shift in how online safety is managed, moving from a reactive to a proactive stance.

At its core, this system leverages advanced machine learning techniques combined with natural language processing. These technologies enable the system to process large amounts of data, identify unhealthy trends in

user behavior, and effec- tively flag questionable content [5]. The careful integration of supervised and unsupervised learning models allows for a nuanced understanding of complex online interactions, en- abling the detection of genuinely harmful exchanges without excessive false positives [6]. Additionally, the system employs facial recognition technology to identify violent visual content, enhancing its ability to address various online threats [10].

An important feature of the system is its ability to define user behavior patterns and preferences. The site can monitor entire conversations, including comments, posts, and even private messages, to identify any signs that someone may be a victim or perpetrator of cyberbullying, or if they might be experiencing mental health issues [8], [9]. For instance, signs like frowning, scowling, or expressing aggression or depres- sive language are flagged as indicators that early intervention and support from a clinician may be needed [7]. Additionally, this functionality plays a crucial role in protecting users and promoting greater acceptance of mental health, making online experiences safer.

Privacy and ethical considerations are vital when designing the lesson plan for ”Guarding Against Online Predators.” The system ensures that all user information is handled with the utmost confidentiality, adhering to global standards for data protection and ethical AI practices [11]. By incorporating privacy by design and encouraging open discussions, trust is built between the platform’s users and the services that help generate revenue in the online market. Furthermore, the platform features an easy-to-use and interactive Graphical User Interface (GUI), along with a backend control panel for designated administrators to filter content and swiftly address any threats or violations of the service’s terms and policies [12].

The potential applications that supporters of the project foresee extend well beyond just personal security. Its features could help businesses monitor overall social media engage- ment, identify customer dissatisfaction, and address such is- sues proactively [6]. The insights gained from the system can also be valuable to educational institutions and policymakers, enabling them to develop better online safety strategies and re- sources [5]. Given the broad range of opportunities it presents, the article “Guarding Against Online Predators” highlights its importance and effectiveness in various contexts.

Considering the diverse and intricate nature of threats, this initiative to provide a comprehensive set of tools addressing different safety aspects redefines the concept. It emphasizes a more proactive and innovative approach to the challenges of today’s digital landscape, ensuring that social networks remain spaces for meaningful communication and genuine connections [3].

The subsequent sections follow this order: Section II exam- ines related works that dealt with comparable problems and approaches regarding online safety. A comparison study in Section III demonstrates both advantages and weaknesses of existing methods. Section IV details how the proposed system was fully implemented together with its essential features. The final section includes the conclusion, it summarizes the essential contributions of this research project and suggests possible improvements for future work.

RELATED WORKS

Monitoring emotions and unsafe content on social media has emerged as a critical challenge in recent years, aimed at making social networks safer and more inclusive. Various approaches have been developed to help identify textual data that may express emotions, sentiment, or instances of cy- berbullying, thereby offering strategies for understanding and managing online threats.

Some studies have explored how to incorporate sentiment, emojis, and other multimodal inputs to enhance detection outcomes. According to the paper [1] proposed a multitask learning approach using Transformer models to identify cy- berbullying behaviors, utilizing both text and emoji inputs. Notably, the blending of

English and Hindi, particularly among younger users, significantly impacts the model's performance. However, issues like data imbalance and the identification of sarcastic text continue to pose challenges for generalizing these models to different contexts.

According to the paper [2], researchers developed a frame- work to identify undesirable behaviors using advanced tools like sentiment analysis, cosine similarity, and conversational graph generation. The program calculates a bullying score, which helps to contextualize online interactions and flag potential instances of cyberbullying. While this approach and its variations show promise for scalability and analyzing conversational characteristics, it primarily relies on textual data and is heavily focused on English news feeds, making it less applicable to multilingual or non-textual media.

According to the paper [3] further advancements in emotion- based detection models can be found, where the generative model has been refined to enhance the explainability of haz- ardous material detection. This framework employs multitask learning as a deep learning technique to improve detection in code-mixed scenarios. However, despite the effectiveness and versatility of these methods in addressing various issues, they come with high computational complexity and depend on annotated datasets, which poses a significant challenge for many applications, particularly due to the limited availability of such data across different regions and languages.

According to the paper [4] research on the improvement of multilingual capabilities has also been explored. This study focused on creating a model that incorporates measures of ag- gression, repetition and intent through fine-tuned transformer models such as m-BERT and MuRIL. Overall, this multilin- gual framework effectively addresses the challenges of code- mixed interactions in languages such as Urdu, Roman Urdu, and English. Nevertheless, despite the model's effectiveness across different cultures and languages, the need for a well- annotated dataset and real-time emotion detection remains a significant barrier to its efficiency across various platforms.

According to the paper [5] explains how emotion detec- tion alongside sentiment analysis serves as a solution for cyberbullying problems. Multiple language models linked to an emotion lexicon system led to improved cyberbullying detection results with increased recall along with F1 score success. Using its ability to read emotional and sentential data in textual messages helps the system accurately evaluate different emotional states. The model operates with specific constraints since it depends on English data and imbalanced datasets that lead to reduced computational speed while also making it difficult to process other languages.

Real-time threatening and sexist language detection has been investigated through the application of YOLO and SSD MobileNetV2 to unstructured data according to the paper [10]. The authors reached high scalability and exact target detection capability through their work on the algorithms themselves. The detection system faced two main disadvantages when dealing with short texts together with increased instances of inaccurate identification results. The processing requirements at every step presented substantial obstacles to system imple- mentation mainly in settings with constrained resources.

According to the paper [7] describes advanced ensemble learning methods which apply AdaBoost with LSTM to- gether with glowworm swarm optimization. High classification performance becomes achievable through the application of sophisticated NLP methods along with embeddings when dealing with Twitter data. The proposed model struggles with generalization because it needs data from particular platforms while its implementation requires complexity and specific platform data.

According to the paper [8] identified how AI-driven models could work alongside an IoT-based online prevention system. The structure uses automatic detection capabilities to check and resolve risky online actions. The combination between cybersecurity systems and machine learning technology boosts the crime

management process by improving decision-making and resource allocations. Recommendation systems face ob- stacles related to privacy issues and implementation problems when used in real-world scenarios along with compatibility difficulties.

These methods highlight both the progress made and the ongoing issues in detecting emotions, sentiment, and negative interactions on social media. As we continue to develop these models, several challenges remain, including the quality of datasets, the ability to adapt to multiple languages, computa- tional costs, and the generalization capabilities of the models.

COMPARISON STUDY

Table 1 provides complete details about artificial intelli- gence techniques used in cyberbullying detection combined with online safety programs. The analysis unifies various studies which utilize NLP and deep learning with sentiment analysis along with BERT, MuRIL, CNN, LSTM and the op- timization technique Glowworm Swarm Optimization (GSO). Each research design uses a distinctive set of detection tools to improve cyberbullying recognition effectiveness. Text- based detection through multitask learning and NLP represents an effective classification technique for code-mixed languages but these methods encounter high processing expenses and small dataset sizes. Research using deep learning methods that include CNN with MobileNetV2 enables automatic image classification but still faces difficulties in working across different media forms. Research employing ensemble learning in combination with optimization approaches yields better detection performance and accuracy though these methods demand substantial computational resources while working exclusively with Twitter data or English-text datasets.

These AI-based strategies operate with several critical ben- efits such as processing content in various languages to- gether with contextual analysis and automatic process detec- tion across multiple social media networks. Several ongoing obstacles exist including language unintentional biases as well as the inability to recognize sarcasm or determine purpose and dependence on top-tier annotated information alongside prob- lems with respect to data privacy norms. Real-time detection systems experience implementation difficulties mostly because they need advanced hardware capacities due to their complex nature.

The research gives an all-encompassing view of current advances in cyberbullying prevention and AI crime prevention through the synthesis of relevant academic studies in a single table. The analysis shows that multiple research approaches advance online security but continues to need work for the development of accessible and accurate systems for various digital domains.The research should focus on developing Machine Learning based solutions which address inclusivity and efficiency to detect cyberbullying in all types of online platforms better.

TABLE I TECHNIQUES, MERITS, AND DEMERITS OF VARIOUS CYBER BULLYING SYSTEMS

TITLE	TECHNIQUES	MERITS	DEMERITS
Emoji, Sentiment and Emotion Aided Cyberbullying Detection in Hinglish Oct 2023[1]	- Multitask Learning - Multimodal Inputs - BERT and MuRIL	- Multitask Learning Approach - Multimodal Framework -Robustness of the Model	- Limited Generalization to Other Languages - Imbalanced Dataset for Sentiment and Emotion

TITLE	TECHNIQUES	MERITS	DEMERITS
BullyNet: Unmasking Cyberbullies on Social Networks April 2021[2]	- Sentiment Analysis - Bullying Score Calculation -Conversation Graph Generation	- Sentiment and Content Anal- ysis - Contextual Analysis -Scalability	- Focus on Text-Based Detection -English-Language Bias -Human Validation Needed
Automatic Cyberbullying Detection: A Mexican Case in High School and Higher Education Students May 2022[3]	- Naive Bayesian Classifier - Random Forest - Logistic Regression - Support Vector Machine -CNN	- Contextual Relevance - Improved Representation of Data -Enhanced Model Comparisons	- Limited Dataset Size - Narrow Social Media Focus - Limited Generalization
Explainable Cyberbullying Detection in Hinglish: A Generative Approach,2024[4]	- Natural Language Processing - Deep Learning - Multitask Learning	- Explainability - Multitask Learning - Effective in Code-Mixed Language	- High Computational Costs - Data Dependency - Limited to Text
Automatic Recognition of Cyberbullying in the Web of Things and social media using Deep Learning Framework,2024[5]	- Deep Learning - BCO -WOT -CNN	- High Accuracy - Adaptability to Different Platforms -Efficient Feature Selection	- High Computational Cost - Dependancy on Quality Data - False Positives and Negatives
A Deep Learning-Based Framework for Offensive Text Detection in Unstructured Data for Heterogeneous Social Media,2023 [6]	- YOLO - SDD MobileNetV2 - OCR	- High Accuracy - Real-Time Detection -Automation - Scalability	- Limited Small Text De- tection - False Positives and neg- atives - High Computational Re-quirements
Multilingual Detection of Cyberbullying in Mixed Urdu, Roman Urdu, and English Social Media Conversations, 2024 [7]	- Fine-tuned m-BERT -MuRIL - Dataset Creation and Annotation	- Addresses cyberbullying across multiple languages. - Incorporates aggression, repetition, and intent measures.	- Limited to specific lan- guages; may not general- ize. - Quality dependent on dataset and annotation. - Challenges in real-

TITLE	TECHNIQUES	MERITS	DEMERITS
		- High detection performance	time detection due to cultural variations.
Cyberbullying Detection Based on Emotion, 2023[8]	- BERT - XLNet - Emotion Detection Model (EDM) - NRC Emotion - AFINN Sentiment Lexicons	- High recall and F1-score improvements. - Validated dataset for emotion detection in cyberbullying.	- Imbalanced dataset challenge. - Requires significant computational resources. - Limited to English datasets.
Ensemble Learning With Tournament Selected GSO Algorithm for Cyberbullying Detection on Social Media, 2023[9]	- LSTM + AdaBoost - Glowworm Swarm Optimization - NLP	- High accuracy - Optimized performance - Effective classification	- High computational cost - Limited to Twitter data - Complex implementa- tion
Cyberbullying Image Classification using Artificial Intelligence for Safer Online Platforms, 2024 [10]	- CNNs - MobileNetV2 - Image Preprocessing Techniques - Flask Framework - TensorFlow and Keras	- Efficiency with Transfer Learning - Scalable Design - Automation of Cyberbullying Detection	- Exclusively Image-Based - Generalization Issues - Limited Multi-Modal Capability - Model Dependency
Detection of Online Humiliation Through Social Media Platforms Using AI Inspired Algorithms, 2023 [11]	- KNN and SGD - NLP - Transfer Learning - Supervised Learning - Big Data Technologies	- High Accuracy - Variety of Algorithms - Prevention of Cyberbullying - Performance Evaluation	- Dataset Limitations - Focus on Textual Data - Generalization - Time-Consuming Algo- rithms
Technological Intervention: Prevention of Crime Using AI and IoT, 2023[12]	- AI - IoT - Machine Learning and Deep Learning - Cybersecurity - Measures	- Enhanced Surveillance and Security - Automation and Resource Optimization - Improved Decision-Making - Broad Applicability	- Privacy and Ethical Con-cerns - Lack of Interoperability - Complexity of Imple- mentation - Cybersecurity Risks

PROPOSED SYSTEM

An effective solution enhances social network safety through a technology system which uses advanced analytics to separate secure from dangerous content. The face recognition module backs the system using the Haarcascade algorithm to stop unauthorized usage of user images. The uploading of pictures triggers a system check which uses stored database images to seek out possible violations. The system automatically alerts users whenever their images get shared improperly therefore granting them time to protect their privacy. Our system uses Haarcascade facial detection which matches uploaded pictures to photos in storage for detecting unauthorized image usage and protecting against identity theft and image abuse.

A. Dataset

Suicide and Depression Detection : The Suicide Watch dataset provided by Kaggle presents suicide-related discussion text data. People share their thoughts emotions along with struggles through online forums which serve as its main source. The dataset serves NLP applications including senti- ment analysis and suicide assessment for mental health moni- toring tasks and risk assessment. The dataset enables scientists to design AI systems which identify mental health risks during early stages of crisis. When using this data researchers need to take ethical steps including privacy protection as well as the management of potential biases and responsible use of information

B. LSTM-Based Comment Toxicity Detection

The system uses NLP and LSTM networks to counter cyberbullying and mental health risks by analyzing text se- quences for abusive actions and emotional distress indicators. Through social text evaluation, it identifies harmful intent and harassment, recognizing cyberbullying and depression via sentiment analysis. The system alerts users and moderators to potential threats, making the online environment safer with enhanced monitoring. Periodic updates keep the system current with new patterns of abusive behavior and incorporate user feedback to optimize its performance, creating a healthy online environment.

C. CNN-Based Inappropriate Content Detection

The system utilizes a Convolutional Neural Network (CNN) to image moderately, identifying bad content through scanning images and videos. The system utilizes CNN technology to scan for unsafe images through analyzing enormous collec- tions of labeled pictures. The system provides binary classi- fication to determine safe or risky content, which provides a safer internet. CNN image moderation is used as an additional feature to text monitoring with higher security. Further, it learns from new information constantly to improve its detection. Through these developments combined, the system guarantees a strong and adaptive method of online protection.

Algorithm 1 LSTM-Based Comment Toxicity Detection Algorithm

1: Input: User comment C, Trained LSTM model MLST M
2: Output: Classification result R (normal or toxic)
3: Procedure:
4: Load dataset D = {(Xi, Yi)} where Xi are messages and Yi are labels
5: Tokenize text using word embeddings
6: Convert text into padded sequences using maximum sen- tence length
7: Define LSTM model: E = Embedding(V, d) ▷ Embedding Layer Ht = LST M(Ht−1, Xt) ▷ LSTM Layer D = Dense(8, activation = ReLU) Yˆ = Dense(1, activation = sigmoid)
8: Train the model using binary cross-entropy loss: $N\ L = -\quad Yi \log(Y\hat{}i) + (1 - Yi) \log(1 - Y\hat{}i) \quad (1)\ N\ i=1$
9: if Prediction Confidence > 0.6 then Classify comment as toxic

10: else Classify comment as normal
11: end if

Algorithm 2 CNN-Based Inappropriate Content Detection Algorithm
1: Input: Image I, Pretrained CNN Model MCNN
2: Output: Classification of inappropriate content
3: Load dataset D containing labeled inappropriate images
4: Preprocess images (resize, normalize)
5: Define CNN architecture:
6: F1 = Conv2D(filters = 32, kernel = 3 × 3, activation = ReLU)
7: P1 = MaxPooling2D(pool size = 2 × 2)
8: F2 = Conv2D(filters = 64, kernel = 3 × 3, activation = ReLU)
9: P2 = MaxPooling2D(pool size = 2 × 2)
10: D = Dense(128, activation = ReLU)
11: Yˆ = Dense(1, activation = sigmoid)
12: Train model using binary cross-entropy loss: N L = – 1 N i=1 hYi log(Yˆi) + (1 – Yi) log(1 – Yˆi)I (2)
13: Feed input image I to MCNN
14: Retrieve classification result based on threshold τ
15: if Yˆ > τ then
16: Output: ”Inappropriate content detected.”
17: else
18: Output: ”No inappropriate content detected.”
19: end if

D. Face Recognition for Unauthorized Image Detection using Haar Cascade

The system design incorporates an alert system for users to inform them about potential hazards found on the platform. Users will get instantaneous alerts whenever the system detects potentially harmful image, text post or other concerning behav- iors. Users receive immediate notifications through the real- time alert system and this enables them to protect themselves and make well-informed choices when using online platforms. Haar Cascade face recognition integration enables the sys- tem to perform efficient unauthorized image detection with strong accuracy levels. This method enables system analysis of facial features which enables database comparison for unau- thorized content identification. The proactive detection system through Haar Cascade protects online safety because it detects problematic images instantly for users and administrators to take appropriate response actions. The detection system joins forces with an alert system to provide users with real-time access to violation reports in order to create a more secure digital environment.

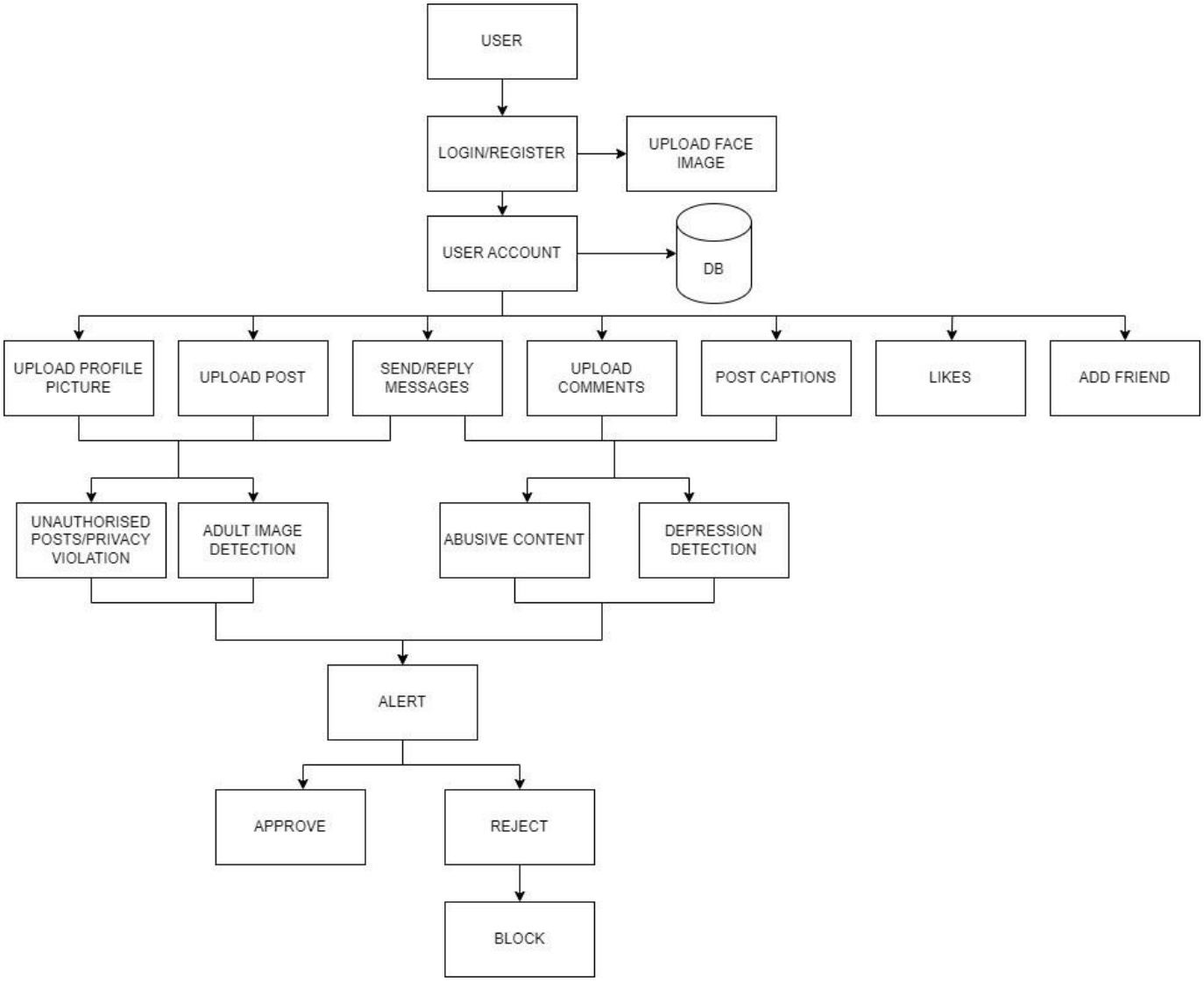


Fig. 1. Architecture diagram of JustPing

Algorithm 3 Haar Cascade-Based Face Recognition for Unauthorized Image Detection

1: Input: Uploaded Image I, Face Database D
2: Output: Face match result
3: Procedure:
4: Convert image I to grayscale
5: Apply Haar Cascade classifier to detect faces:
6: F = haarcascade.detectMultiScale(I)
7: for each detected face region Fi ∈ F do
8: Extract face features using eigenfaces or LBPH
9: for each stored face Fj ∈ D do
10: Compute similarity:
11: Sij = Fi – Fj
12: if Sij < τ (Predefined Threshold) then
13: Notify user about unauthorized image usage
14: end if
15: end for

16: end for
17: Save processed image and return status

It incorporates Haarcascade facial recognition, LSTM-based NLP, and CNN image moderation in offering homogenized digital security. Regular updates and user feedback facilitate threat detection and content moderation, making the system adaptable against up-and-coming threats. Suspicious content can be reported by users, allowing AI-human moderation for enhanced security. Blending automation with user engagement, the system promotes a safe and accountable online culture.

PERFORMANCE ANALYSIS

The evaluation of machine learning detection algorithms for cyberbullying and adult content and face recognition utilizes key metrics that encompass accuracy, precision, recall and F1- score for evaluation purposes. The evaluation takes place using standard datasets for each category.

A. Cyberbullying Detection

The model proves its capability in detecting cyberbullying from other non-bullying content. The performance data can be found in Table II.

TABLE II CYBERBULLYING DETECTION PERFORMANCE METRICS

Sentiment	Precision	Recall	F1-Score	Support
Non-Bullying	0.89	0.91	0.90	2000
Cyberbullying	0.85	0.83	0.84	1500
Overall Accuracy	**0.88**			

The model reaches an 88% accuracy in its cyberbullying detection tasks while maintaining strong detection capabilities. The model performance metrics including precision and recall and F1-score for both categories where balanced performance is displayed.

B. Sentiment Analysis

Text entered in the model leads to sentiment classification as positive or negative or neutral. The presented performance results are displayed in Table III.

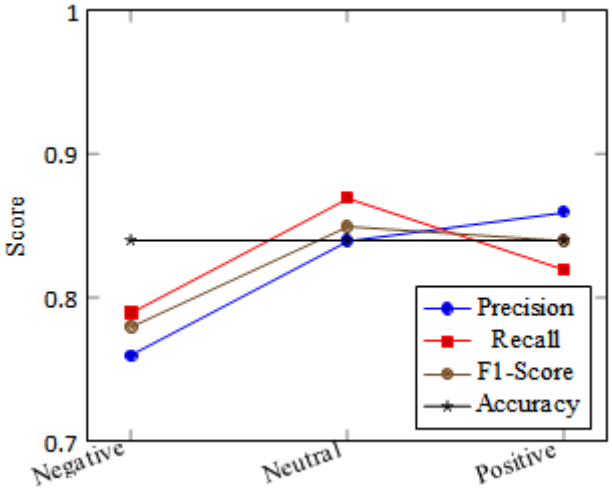
TABLE III SENTIMENT ANALYSIS PERFORMANCE METRICS

Sentiment	Precision	Recall	F1-Score	Support
Negative	0.76	0.79	0.78	246
Neutral	0.84	0.87	0.85	714
Positive	0.86	0.82	0.84	709
Overall Accuracy	**0.84**			

The identification of sentiment categories achieves an 84% accuracy score with F1-scores higher than 0.78. The model reaches its optimal discrimination results for senti- ment categories when it identifies neutral texts with the highest success ratio according to Figure 3. The model demonstrates high efficiency in separating different senti- ments because its recall and precision results show minimal confusion between categories. Model generalization occurs through the balanced F1-scores obtained across sentiment categories which indicates

stability when processing different input data points. This strong performance proves the relia- bility of the sentiment analysis model which it runs in real- world platforms such as social media monitoring along with customer feedback analysis and online content moderation.

Sentiment Analysis Performance



C. Adult Content Detection

The model differentiates material content suitable for adults from content not intended for adult populations. Table IV displays the obtained performance data.

TABLE IV ADULT CONTENT DETECTION PERFORMANCE METRICS

Content Type	Precision	Recall	F1-Score	Support
Non-Adult	0.95	0.94	0.94	3000
Adult Content	0.97	0.96	0.96	2000
Overall Accuracy	**0.96**			

The model demonstrates 96% precision in its operation to identify adult content. .

D. Face Recognition

The model functions to identify facial images as either recognized or unidentified. The performance results can be found in Table V.

TABLE V FACE RECOGNITION PERFORMANCE METRICS

Recognition	Precision	Recall	F1-Score	Support
Known Face	0.93	0.92	0.92	2500
Unknown Face	0.91	0.90	0.91	1500
Overall Accuracy	**0.92**			

The model reaches 92% accuracy during face recognition procedures to differentiate known from unknown faces. The model reveals its successful face recognition abilities through precision, recall and F1-score exhibit minimal interferences.

E. Comparison and Observations

- The detection model reaches an evaluation accuracy of 88% while maintaining stable precision alongside recall for identifying cyberbullying encounters.
- The sentiment classification model shows consistent per- formance in all categories by reaching an accuracy level of 84%.
- The detection of adult content reaches a 96% accuracy level.
- The model achieves 92% accuracy that enables it to separate recognized faces from unrecognized ones.
- The models in their evaluation show superior accuracy for multiple classification duties while proving useful for practical implementation.

CONCLUSION

The development of a Social Media Crime Prevention App showcases how artificial intelligence and machine learning can help reduce cyber threats and promote safer online in- teractions. By incorporating real-time monitoring, automated threat detection, and user reporting features, this proposed system effectively identifies and prevents various types of cybercrime, such as cyberbullying, harassment, and fraud.The research emphasizes the importance of utilizing advanced AI algorithms, natural language processing (NLP), and image recognition techniques to analyze user interactions and detect suspicious behavior. The model has undergone rigorous testing through multiple iterations, demonstrating high accuracy in identifying potential threats while keeping false positives to a minimum.However, some limitations remain. The system primarily focuses on recognizing predefined patterns and may have difficulty with new threats that develop over time. Addi- tionally, adding multilingual support and enhancing context- aware analysis could improve the system's overall effective- ness. Future efforts will aim to expand the dataset, refine deep learning models, and incorporate blockchain-based security measures to bolster data integrity and user privacy.In summary, this study establishes a foundation for creating an AI-driven social media crime prevention system. With ongoing advance- ments in AI and cybersecurity, the proposed solution can be further improved to foster a safer and more responsible digital landscape.

REFERENCES

[1]. K. Maity, S. Saha and P. Bhattacharyya, "Emoji, Sentiment and Emotion Aided Cyberbullying Detection in Hinglish," in IEEE Transactions on Computational Social Systems, vol. 10, no. 5, pp. 2411-2420, Oct. 2023, doi: 10.1109/TCSS.2022.3183046.

[2]. A. S. Srinath, H. Johnson, G. G. Dagher and M. Long, "BullyNet: Unmasking Cyberbullies on Social Networks," in IEEE Transactions on Computational Social Systems, vol. 8, no. 2, pp. 332-344, April 2021, doi: 10.1109/TCSS.2021.3049232.

[3]. K. I. Arce-Ruelas, O. Alvarez-Xochihua, L. Pellegrin, L. Cardoza- Avendan˜o and J. A´ . Gonza´lez-Fraga, "Automatic Cyberbullying Detec- tion: a Mexican case in High School and Higher Education students," in IEEE Latin America Transactions, vol. 20, no. 5, pp. 770-779, May 2022, doi: 10.1109/TLA.2022.9693561.

[4]. K. Maity, R. Jain, P. Jha and S. Saha, "Explainable Cyberbullying Detection in Hinglish: A Generative Approach," in IEEE Transactions on Computational Social Systems, vol. 11, no. 3, pp. 3338-3347, June 2024, doi: 10.1109/TCSS.2023.3333675.

[5]. F. N. Al-Wesabia et al., ”Automatic Recognition of Cyberbullying in the Web of Things and social media using Deep Learning Framework,” in IEEE Transactions on Big Data, doi: 10.1109/TBDATA.2024.3409939.

[6]. J. Bacha, F. Ullah, J. Khan, A. W. Sardar and S. Lee, ”A Deep Learning- Based Framework for Offensive Text Detection in Unstructured Data for Heterogeneous Social Media,” in IEEE Access, vol. 11, pp. 124484- 124498, 2023, doi: 10.1109/ACCESS.2023.3330081.

[7]. F. Razi and N. Ejaz, ”Multilingual Detection of Cyberbullying in Mixed Urdu, Roman Urdu, and English Social Media Conversations,” in IEEE Access, vol. 12, pp. 105201-105210, 2024, doi: 10.1109/AC-CESS.2024.3432908.

[8]. M. Al-Hashedi, L. -K. Soon, H. -N. Goh, A. H. L. Lim and E. -G. Siew, ”Cyberbullying Detection Based on Emotion,” in IEEE Access, vol. 11, pp. 53907-53918, 2023, doi: 10.1109/ACCESS.2023.3280556.

[9]. R. Daniel et al., ”Ensemble Learning With Tournament Selected Glow- worm Swarm Optimization Algorithm for Cyberbullying Detection on Social Media,” in IEEE Access, vol. 11, pp. 123392-123400, 2023, doi: 10.1109/ACCESS.2023.3326948

[10]. M. V. Krishna, R. Asish Verma and P. Kirubanantham, ”Cyberbully- ing Image Classification using Artificial Intelligence for Safer Online Platform,” 2024 4th International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, 2024, pp. 468-474, doi: 10.1109/ICPCSN62568.2024.00079.

[11]. A. Bhatia, A. Kumar, Neetu, A. Kumar, S. Sachi and S. Ku- mar, ”Detection of Online Humiliation Through Social Media Plat- forms Using AI Inspired Algorithms,” 2023 3rd International Con- ference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2023, pp. 364-367, doi: 10.1109/IC- TACS59847.2023.10390175.

[12]. S. Singh, N. Yamsani, V. Uniyal, M. Sahu, S. Pandey and A. Gehlot, ”Technological Intervention: Prevention of Crime Using AI and IoT,” 2023 IEEE World Conference on Applied Intelligence and Computing (AIC), Sonbhadra, India, 2023, pp. 778-782, doi: 10.1109/AIC57670.2023.10263817