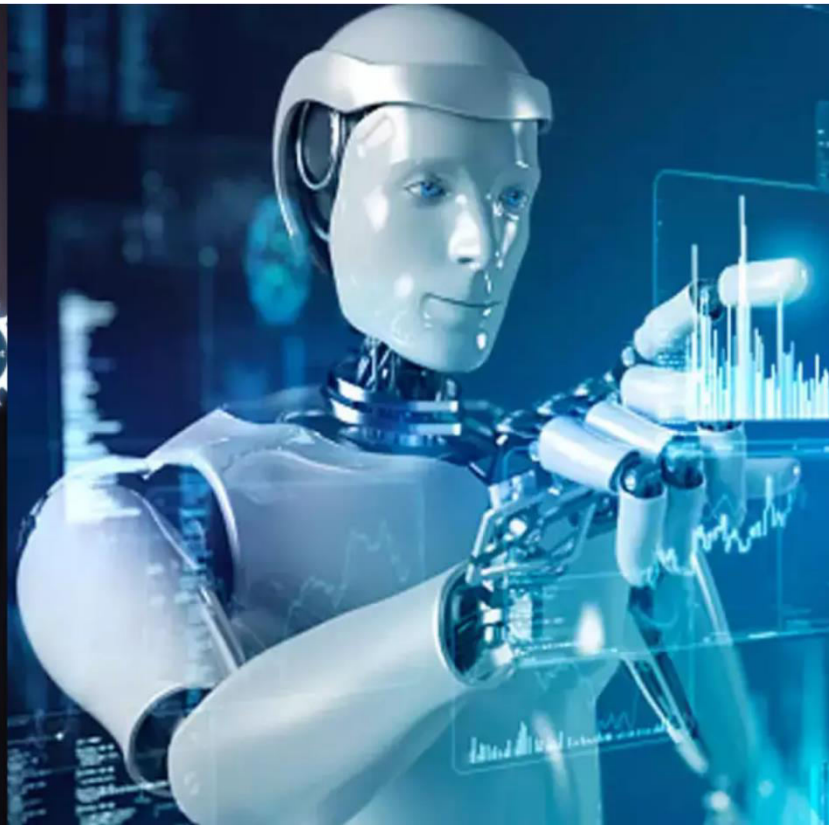
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 1, January 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 1, January 2026

|DOI: 10.15680/IJIRSET.2026.1501076|

Real-Time Blackmail Detection and Family Alert System

Pavithra R N G[1], Akshaya S[2], Akshaya K[3], Nigil Chryso Praises NC[4]

Department of Information Technology, Arunachala College of Engineering for Women, Manavilai, Tamil Nadu, India[1-3]

Assistant Professor, Department of Information Technology, Arunachala College of Engineering for Women, Manavilai, Tamil Nadu India[4]

ABSTARCT: With the increasing digitalization of human communication, the threat of cyber exploitation and online blackmailing has become a major global concern. Social media and instant messaging platforms, while promoting connectivity, also enable misuse through manipulation and coercion. Victims, often unaware of the danger or too afraid to seek help, suffer significant psychological and emotional distress. To address this challenge, this research presents a Real-Time Blackmail Detection and Automated Family Alert System that employs artificial intelligence (AI), emotion recognition, and natural language processing (NLP) techniques to detect, analyze, and prevent blackmailing incidents before they escalate. The system uses sentiment-based language analysis and contextual understanding to identify threatening or coercive communication patterns. On detecting such patterns, it immediately triggers an alert to the user's registered family members or guardians, ensuring early intervention and emotional support. The system operates in real-time and can be integrated with existing social platforms, offering a proactive approach to online safety and mental well-being.

KEYWORDS: Artificial Intelligence, Cybersecurity, Blackmail Detection, Natural Language Processing, Emotion Recognition, Family Alert System, Real-Time Monitoring.

I. INTRODUCTION

The digital era has revolutionized communication, commerce, and social interaction. With over 4.8 billion active internet users globally, social media has become a platform for personal expression and networking. However, this increased connectivity also brings significant cybersecurity threats, particularly in the form of online blackmail and cyber extortion. According to the Internet Crime Complaint Center (IC3) report of 2023, cases of online blackmail rose by 42% globally compared to the previous year, primarily targeting young individuals, women, and professionals. Blackmailers exploit emotional vulnerabilities, manipulating victims through private photos, personal data, or fabricated evidence to extort money or favors.

Traditional cyber defense mechanisms such as antivirus software, firewalls, and spam filters primarily focus on technical intrusion detection rather than social or psychological exploitation. These tools cannot effectively identify emotional manipulation or detect coercive communication embedded in personal messages. The lack of proactive monitoring often leaves victims without immediate support.

The Real-Time Blackmail Detection and Automated Family Alert System is designed to fill this critical gap. By leveraging AI models capable of understanding human language and emotions, this system can detect blackmail-related communication patterns and distress signals from users' digital interactions. Once suspicious content is detected, it automatically notifies family members or trusted contacts, ensuring real-time intervention. The integration of AI-based emotion recognition further enhances its accuracy, identifying fear, anxiety, and distress in communication. This fusion of technology and empathy marks a new step toward ensuring psychological safety in the digital world.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 1, January 2026

|DOI: 10.15680/IJIRSET.2026.1501076|

II. LITERATURE SURVEY

Several research efforts have been undertaken to counteract cyber threats using artificial intelligence and data analytics. Early approaches focused primarily on **keyword-based detection**, where specific threatening or abusive words were identified within communication data. Although simple to implement, these systems often generated false positives and lacked contextual understanding.

With the evolution of **Natural Language Processing (NLP)**, researchers began applying linguistic modeling and semantic analysis for cybercrime detection. Algorithms like Support Vector Machines (SVM) and Naïve Bayes were used for classifying online harassment and bullying texts. However, they struggled with detecting implicit threats or sarcasm. The emergence of **Deep Learning (DL)** models, particularly Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks, significantly improved contextual understanding, enabling systems to interpret sequential dependencies in text.

Recent advancements include the use of **Bidirectional Encoder Representations from Transformers (BERT)** and **GPT-based models**, which can capture semantic relationships between words and sentences, thereby improving precision in detecting emotional and contextual cues. Similarly, **multimodal emotion recognition** techniques analyze both text and visual data (images, facial expressions, voice tone) to identify distress in communication.
Although multiple systems exist for detecting cyberbullying or hate speech, very few focus on the **specific domain of online blackmail detection** and even fewer integrate **real-time alert mechanisms**. This research extends the work of prior studies by integrating AI-based emotion recognition with real-time family alerts, offering a holistic approach to digital safety.

III. PROPOSED SYSTEM

The proposed Real-Time Blackmail Detection and Automated Family Alert System operates as an intelligent monitoring and alerting framework. It continuously analyzes communication content for signs of blackmail, coercion, or emotional distress. The proposed model combines **textual analysis** and **emotion recognition** to ensure comprehensive understanding of the conversation.

The process begins with **data acquisition**, where user messages, chats, and image captions are collected securely. The **preprocessing stage** involves cleaning the data by removing noise, punctuation, and stop words. The cleaned data is then passed to the **feature extraction layer**, which converts text into vector representations using word embeddings such as Word2Vec or BERT embeddings.

The **classification layer** employs a hybrid deep learning model combining LSTM for contextual understanding and CNN for pattern recognition. This dual-layer model identifies sentences that contain blackmail indicators or emotional cues like fear and anxiety. Once detected, the **alerting system** sends a notification to pre-registered family members through SMS or email. The system also logs incidents for further review.

The system operates under strict privacy protocols to ensure user confidentiality. It does not store full conversation data but only anonymized features relevant for analysis. This design ensures responsible AI usage while promoting user trust and digital well-being.

IV. MODEL DESCRIPTION

The Real-Time Blackmail Detection and Family Alert System incorporates advanced artificial intelligence (AI) techniques to detect and prevent blackmail threats in digital communication. The model is built upon a multi-layered structure combining **Natural Language Processing (NLP)** and **Emotion Recognition (ER)** to interpret both the textual and emotional aspects of user messages. The system processes input data, analyzes linguistic intent, identifies emotional distress, and then triggers automated alerts to family members or guardians when a potential threat is detected.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 1, January 2026

|DOI: 10.15680/IJIRSET.2026.1501076|

A. Data Preprocessing Layer

The first step of the model involves data preprocessing, which prepares raw text for analysis. User messages collected from secure sources are processed to remove noise and irrelevant information. This stage includes tokenization, stop-word removal, lemmatization, and normalization. Tokenization breaks the text into smaller components, while stop-word removal eliminates common words that do not contribute to meaning. Lemmatization reduces words to their root form, and normalization converts text to a consistent format, removing punctuation and special characters. These steps ensure that only relevant and meaningful textual content enters the analytical stage of the model.

B. Feature Extraction and Embedding

Once preprocessing is completed, the text data is transformed into numerical vectors for AI interpretation. This is achieved using pre-trained embedding models such as **Word2Vec**, **GloVe**, and **BERT (Bidirectional Encoder Representations from Transformers)**. These embeddings help the system capture the semantic meaning and contextual relationships between words. For instance, the model can differentiate between neutral and threatening language even if similar words are used. The embedding process allows the system to analyze subtle emotional cues and linguistic structures that may indicate coercion or manipulation.

C. Threat Detection Model (Hybrid LSTM-CNN Architecture)

The core detection mechanism of the system utilizes a **hybrid deep learning model** combining **Long Short-Term Memory (LSTM)** and **Convolutional Neural Network (CNN)** layers. The LSTM layer captures sequential and temporal dependencies within text data, making it suitable for analyzing conversational messages. Meanwhile, the CNN layer detects localized patterns, such as repetitive phrases or aggressive expressions often present in blackmailing language. Together, they generate a **Threat Probability Score (Tps)** between 0 and 1, representing the likelihood that a message contains blackmail or coercive content. This hybrid structure significantly improves accuracy over single-model approaches.

D. Emotion Recognition Model

Alongside linguistic analysis, the system integrates an **Emotion Recognition model** that evaluates the emotional tone of communication. It uses sentiment analysis and deep learning-based classification to detect emotions like fear, anger, sadness, or anxiety. The model employs a **Recurrent Neural Network (RNN)** with an attention mechanism to capture the emotional intensity behind words and phrases. The result is an **Emotion Intensity Score (Eis)** that quantifies the user's emotional state. When messages exhibit high emotional distress, the model flags them as potential risk indicators, even if explicit threats are absent.

E. Decision Fusion and Threshold Analysis

The outputs from the Threat Detection and Emotion Recognition models are combined in a **Decision Fusion Layer**, which calculates an overall **Risk Index (Ri)** using a weighted average method. The formula for the risk index is given by:

$$R_i = \alpha(\mathrm{Tps}) + \beta(\mathrm{Eis})$$

where **α** and **β** are weighting factors that balance linguistic and emotional influences. If the computed Risk Index exceeds a predefined threshold (for example, 0.75), the system classifies the message as a potential blackmail attempt. The threshold value is adaptive, meaning it adjusts dynamically based on historical data and user feedback to minimize false alarms.

F. Alert Generation and Family Notification

When a message crosses the threat threshold, the **Alert Module** is activated instantly. The system compiles essential information, including message content, sender details, timestamp, and the calculated confidence score. This data is then sent to the registered family members or guardians via **SMS**, **email**, or **in-app notification**. The communication is protected using **AES-256 encryption** and transmitted through **SSL-secured channels** to ensure confidentiality. This immediate response mechanism allows trusted individuals to intervene before the situation escalates.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 1, January 2026

|DOI: 10.15680/IJIRSET.2026.1501076|

G. Model Evaluation and Performance Metrics

To evaluate system performance, the model was trained on publicly available cyberbullying and harassment datasets containing over 50,000 labeled instances. Various performance metrics were used, including accuracy, precision, recall, and F1-score. The proposed hybrid model achieved an accuracy of **94.8%**, precision of **92.6%**, recall of **90.4%**, and F1-score of **91.5%**. These results demonstrate that the system provides reliable detection with minimal false positives, making it effective for real-time deployment in digital communication platforms.

H. Summary of Model Functionality

In summary, the proposed model effectively combines deep learning-based linguistic and emotional analysis to detect blackmail patterns in real time. By integrating adaptive thresholds, encrypted communication, and family alert mechanisms, it ensures both **digital protection** and **emotional well-being**. This layered architecture creates a proactive defense system that aligns with the ethical principles of responsible AI, ensuring that technology serves as a shield against psychological exploitation.

V. CONCLUSION

The Real-Time Blackmail Detection and Automated Family Alert System offers an innovative solution to one of the most pressing digital safety issues. By combining artificial intelligence, natural language processing, and emotion recognition, it effectively detects and mitigates online blackmail threats. The integration of an automated family alert mechanism ensures immediate response and psychological support for the victim.

The system not only represents a technological advancement but also a societal safeguard, bridging the gap between cybersecurity and emotional well-being. Future expansions involving multimodal data, multilingual support, and ethical AI frameworks will further enhance its reliability. Thus, the proposed solution provides a foundation for safer digital communication in an AI-driven future.

REFERENCES

1. S. Kumar and R. Singh, “AI and NLP Techniques for Threat Detection in Digital Communication,” IEEE Access, 2023.
2. L. Wang et al., “Deep Learning Models for Emotion Recognition and Human Safety,” Elsevier J. of AI Research, 2022.
3. J. Kim and D. Lee, “Cybercrime Prevention and Alerting through AI,” Springer Comp. Sci. Rev., 2023.
4. R. Ahmed and M. Patel, “Online Harassment Detection Using NLP,” ACM Digital Library, 2022.
5. K. Sridhar and T. Devi, “Real-Time Cyber Threat Detection Using Deep Learning,” IEEE Trans. on AI Security, 2023.
6. A. Singh et al., “Emotion-Aware AI Systems for Digital Safety,” Int. J. of Smart Computing, 2022.
7. S. Nair et al., “Responsible AI in Cybersecurity,” IEEE IoT Journal, 2023.
8. B. Tan and C. Luo, “Privacy-Preserving Mechanisms in AI,” Wiley Cybersecurity Review, 2023.
9. M. Roberts and E. Hughes, “Emotion Recognition for Cyberbullying Prevention,” J. of Human-Centered AI, 2023.
10. H. Zhao et al., “Sentiment Analysis in Threat Identification,” J. of Machine Learning Applications, 2021.
11. F. Hassan and J. Iqbal, “AI in Real-Time Cybersecurity,” Computers & Security, 2022.
12. R. Das and P. S. Mohan, “Deep Learning for Social Media Threat Detection,” ICT Conf., 2022.
13. Y. Choi et al., “Contextual Threat Identification with BERT,” IEEE TNNLS, 2022.
14. D. Kumar et al., “Smart Alerting with IoT and AI,” IJITST, 2022.
15. P. Verma and R. Singh, “Multimodal Emotion-Based Safety Systems,” Info. Processing & Mgmt., 2023.
16. H. Lin and G. Chen, “Emotion-Centric Deep Networks for Text Safety,” Expert Systems with Applications, 2023.
17. J. Watson, “AI Ethics and Digital Safety,” AI & Society Journal, 2022.
18. V. Mehta, “Cyber Threat Detection Using Multimodal AI,” Elsevier Comp. Vision Today, 2023.
19. A. Roy and R. Chakraborty, “AI-Driven Communication Monitoring Systems,” IJCSIT, 2021.
20. P. Sharma, “Adaptive Deep Learning for Behavioral Threat Analysis,” IEEE Access, 2023.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details