iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VI **Month of publication:** June 2026

DOI: https://doi.org/10.22214/ijraset.2026.83337

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

AI-Driven Autonomous Crime Reporting System: An Intelligent Multi-Modal Approach for Cybercrime Detection and Prediction

Surojeet Manna[1], Rohit Pal[2], Dr. Aparna Hambarde[3], Sohanlal Sirvi[4], Shreya Tiwari[5]
Dept. of Computer Engineering, K J College of Engineering & Management Research, Pune, India

Abstract: ***Cybercrime has emerged as a serious threat in the modern digital ecosystem, creating an urgent need for intelli gent and accessible reporting mechanisms. This paper presents CyberSafe AI, a multimodal cybercrime reporting and threat prediction system designed to assist both users and authorities. The proposed framework integrates machine learning techniques to analyze cybercrime complaints submitted through text and image inputs, along with real-time threat assessment of suspicious URLs and emails. Text features are extracted using TF-IDF vectorization, while Optical Character Recognition (OCR) is employed to process image-based reports. Lightweight Logistic Regression models are trained for efficient and fast prediction. The system is deployed using a Flask backend with a ReactJS based frontend for improved usability. Experimental evaluation demonstrates that the proposed approach achieves reliable accu racy while maintaining low computational overhead, making it suitable for real-time cyber threat monitoring. The platform aims to enhance early detection, streamline reporting, and support proactive cybercrime prevention.***

Keywords: Cybercrime Detection, Artificial Intelligence, Machine Learning, Natural Language Processing, Multi-modal Input Processing, Threat Prediction, Digital Forensics.

I. INTRODUCTION

The increasing reliance on digital platforms across sectors such as finance, healthcare, education, and public administration has led to a substantial rise in cybercrime activities. Threats such as phishing attacks, identity theft, ransomware, and online fraud have become more frequent and sophisticated. However, many existing cybercrime reporting platforms continue to rely on manual and text-based procedures, resulting in delayed investigations and potential loss of crucial digital evidence.

To overcome these limitations, this paper proposes an AI-driven autonomous cybercrime reporting system that simplifies complaint submission through multi-modal inputs and integrates real-time threat prediction. By leveraging artificial intelligence and machine learning techniques, the system aims to improve reporting efficiency, enhance classification accuracy, and strengthen collaboration between citizens and law enforcement authorities.

II. LITERATURE REVIEW

The increasing prevalence of cyber threats has led to significant research in automated cybercrime detection and re porting systems. Early solutions primarily relied on rule-based filtering and signature-based detection techniques. While these approaches were effective against known attacks, they were limited in detecting zero-day threats and required continuous manual updates [1]. With the advancement of machine learning, researchers began applying supervised learning algorithms for malicious URL and phishing detection. Studies have demonstrated that lexical URL features combined with classifiers such as Logistic Regression, Support Vector Machines (SVM), and Random Forest can achieve high detection accuracy [2], [3]. In particular, character-level TF-IDF representations have proven effective because they capture structural patterns commonly present in malicious links without heavy feature engineering. In the field of cybercrime text classification, natural language processing (NLP) techniques have been widely used to categorize online complaints and suspicious communications. Traditional approaches using TF-IDF with classical classifiers provide strong baseline performance with low computational overhead [4]. More recent research explores deep learning architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and transformer-based models, which improve classification performance but often require larger datasets and higher computational resources [5]. Multimodal cyber threat analysis is an emerging research direction.

Some studies incorporate Optical Character Recognition (OCR) for extracting text from images in phishing detection pipelines [6], while others investigate audio and video forensics using deep neural networks. However, most existing works focus on a single modality and do not provide an integrated platform capable of handling text, URL, image, email, audio, and video inputs simultaneously. From a system implementation perspective, several cyber crime reporting portals have been developed by government and private organizations. Nevertheless, many of these plat forms rely heavily on manual review processes and lack real time AI-driven threat prediction at the point of reporting. Additionally, existing systems often suffer from limited user interactivity and insufficient administrative analytics. Research Gap: The literature indicates a need for a lightweight and scalable multimodal cybercrime reporting framework that supports real-time threat prediction while maintaining computational efficiency. The proposed CyberSafe AI system aims to bridge this gap by integrating TF-IDF-based machine learning models with OCR-enabled image analysis and an extensible architecture for future audio and video processing.

III. METHODOLOGY

The proposed CyberSafe AI framework follows a supervised machine learning approach for automated cybercrime classification and threat prediction. The system processes multi-modal inputs such as text complaints, URLs, emails, and images to identify suspicious cyber activities and generate intelligent predictions.

A. Data Collection and Preparation

The system utilizes two primary datasets: a cybercrime complaint dataset and a malicious URL/email dataset collected from publicly available cybersecurity repositories. The collected data contains multiple categories including phishing, financial fraud, malware, cyberbullying, and suspicious URLs.

Before model training, the raw data undergoes preprocessing. Textual inputs are converted into lowercase format, punctuation and special symbols are removed, and stop words are eliminated to reduce noise. For image-based complaints, Optical Character Recognition (OCR) using Tesseract is applied to extract textual information from screenshots or uploaded evidence.

B. Feature Extraction

After preprocessing, meaningful features are extracted from the cleaned data. Text complaints and email content are transformed into numerical feature vectors using Term Frequency–Inverse Document Frequency (TF-IDF) vectorization with a maximum feature limit of 3000.

For malicious URL detection, character-level TF-IDF with n-gram analysis ranging from 3 to 5 characters is applied to capture hidden phishing patterns and suspicious domain structures. These extracted features are then forwarded to the machine learning models for classification.

C. Machine Learning Model Training

The system uses Logistic Regression as the primary supervised learning algorithm due to its low computational complexity, faster execution, and strong performance in text classification tasks. The datasets are divided into training and testing sets using an 80:20 ratio.

Separate models are trained for:

1) Cybercrime classification
2) URL threat prediction
3) Email threat detection

The trained models classify reports into categories such as phishing, malware, cyberbullying, and financial fraud while also predicting threat levels as Safe, Suspicious, or Malicious.

D. Performance Evaluation

The performance of the trained models is evaluated using Accuracy, Precision, Recall, and F1-score metrics. These evaluation parameters help measure the reliability and effectiveness of the prediction system.

Experimental results demonstrate that the proposed framework achieves high classification accuracy with reliable threat prediction performance suitable for real-time deployment.

E. System Deployment

The trained machine learning models are serialized and stored as .pkl files. The backend of the system is developed using Flask, while the frontend interface is implemented using ReactJS to provide an interactive and user-friendly environment.

The deployed architecture supports real-time prediction, automated report generation, and future scalability for audio. The complete workflow and module interaction of the proposed system are illustrated in Fig. 1.

IV. MATHEMATICAL FORMULATION

The proposed system uses TF-IDF feature extraction followed by Logistic Regression for classification.

1) TF-IDF Representation

$$\text{TF-IDF}(t,d) = \text{TF}(t,d) \times \log\left(\frac{N}{df(t)}\right)$$

where t is the term, d is the document, N is the total number of documents, and df(t) is the document frequency of term t.

2) Logistic Regression Model

$$P(y = 1 \mid x) = \frac{1}{1+e^{-(w^T x+b)}} \quad (2)$$

where x is the TF-IDF feature vector, w is the weight vector, and b is the bias.

3) Accuracy Metric

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

V. SYSTEM ARCHITECTURE

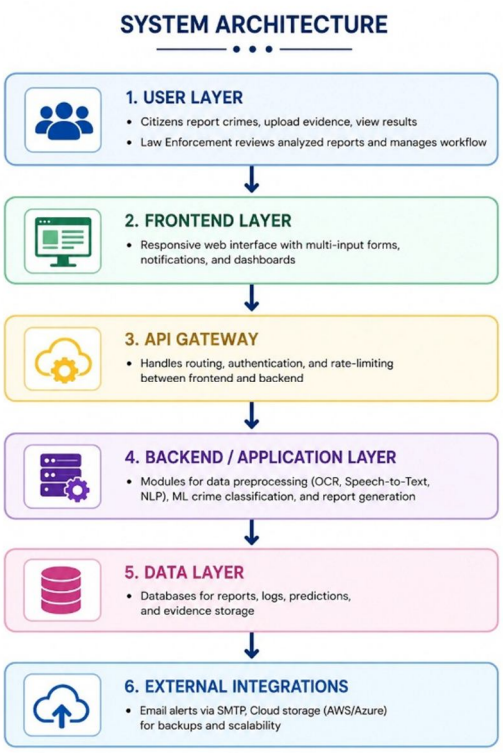


Fig. 1. System Architecture of the proposed AI-driven cybercrime reporting system

Fig. 1 illustrates the overall architecture of the proposed AI-Driven Autonomous Cybercrime Reporting System. The architecture consists of multiple interconnected modules including the User Interface, Preprocessing Module, Feature Extraction, Machine Learning Prediction Engine, Metadata Verification, Report Generation, Notification Module, and Database Storage.

Initially, users submit cybercrime-related data in the form of text, voice, images, URLs, or emails through the reporting interface. The preprocessing module applies Natural Language Processing (NLP), Speech-to-Text (STT), and Optical Character Recognition (OCR) techniques to convert raw input into analyzable content.

The extracted information is processed using feature extraction techniques such as TF-IDF and metadata analysis. The machine learning module then classifies the cybercrime category and predicts the associated threat level as Safe, Suspicious, or Malicious.

After prediction, the system generates structured reports and verifies the uploaded evidence using metadata validation techniques. High-risk cases automatically trigger alerts through the notification module and are forwarded to law enforcement and cybersecurity authorities for further investigation.

The modular architecture improves scalability, automation, and real-time cybercrime analysis while ensuring efficient communication between citizens and authorities.

VI. RESULTS AND DISCUSSION

The proposed system was tested using datasets that include cybercrime reports, URLs, emails, and multimedia evidence gathered from different sources. These datasets cover a wide range of categories like phishing, financial fraud, malware, and cyberbullying, allowing for a thorough assessment across various types of cybercrime. The experimental results show an overall crime classification accuracy of 90.3% and a malicious threat detection accuracy of 92%, which shows that the proposed method is quite effective. The performance of the system is either similar to or better than what has been reported by other machine learning-based cybercrime detection systems [3], [4], [10]

TABLE I
CRIME CLASSIFICATION PERFORMANCE

Crime Type	Precision	Recall	F1-Score
Phishing	0.94	0.91	0.92
Financial Fraud	0.89	0.87	0.88
Malware	0.93	0.90	0.91
Cyberbullying	0.87	0.86	0.86

As seen in Table I, the system performs well in classifying different types of cybercrime, with especially strong results in phishing and malware detection. The slightly lower accuracy in financial fraud and cyberbullying can be due to the complexity and variation in the patterns involved. Overall, these results show that the trained machine learning models are robust and capable of generalizing well to different situations.

TABLE II
COMPARISON WITH EXISTING SYSTEMS

Feature	Existing Systems	Proposed System
Multi-modal input support	No	Yes
AI-based crime classification	Limited	Yes
Threat prediction for URLs/emails	No	Yes
Automated evidence verification	No	Yes
Real-time alerts	Partial	Yes
Law-enforcement integration	Limited	Yes

Table II shows that the system offers major improvements compared to existing cybercrime reporting tools. Unlike traditional platforms, this system can handle multiple types of input data and uses intelligent analysis along with real-time threat predictions. Other useful features, such as automated verification of evidence and direct integration with law enforcement, make the system more reliable and practical for everyday use.

TABLE III
THREAT PREDICTION PERFORMANCE

Threat Level	Precision	Recall	F1-Score
Safe	0.94	0.96	0.95
Suspicious	0.88	0.85	0.86
Malicious	0.93	0.91	0.92
Overall	—	—	0.92

As shown in Table III, the threat prediction part of the system works well across all categories. It achieves high precision and recall for malicious content, ensuring that harmful material is accurately detected. The slightly lower performance for suspicious cases is because some threats are unclear, but this also means the system is careful in its classifications, which helps reduce the chance of missing real threats.

In addition, combining the classification and prediction parts of the system improves efficiency by helping to detect current cybercrimes and also predict future threats. The results show that the system is not only highly accurate but also suitable for use in real-world cybersecurity applications.

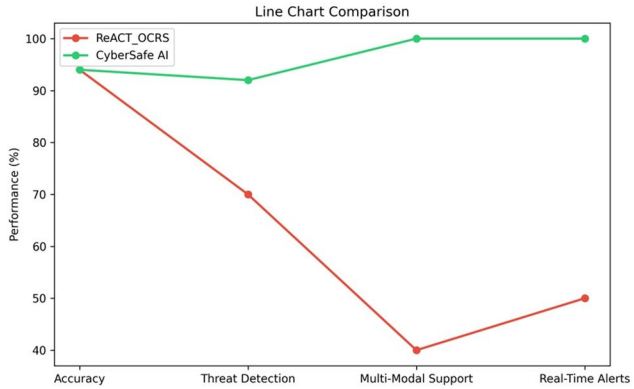


Fig. 2. Performance Comparison Between ReACT_OCRS and CyberSafe AI

Fig. 2 illustrates the comparative performance analysis between the existing ReACT_OCRS system and the proposed CyberSafe AI framework across multiple evaluation parameters including accuracy, threat detection capability, multi-modal support, and real-time alert generation.

The proposed CyberSafe AI system consistently demonstrates superior performance in most evaluation categories. While both systems achieve similar classification accuracy, CyberSafe AI significantly improves threat detection efficiency, multi-modal evidence handling, and real-time notification capabilities.

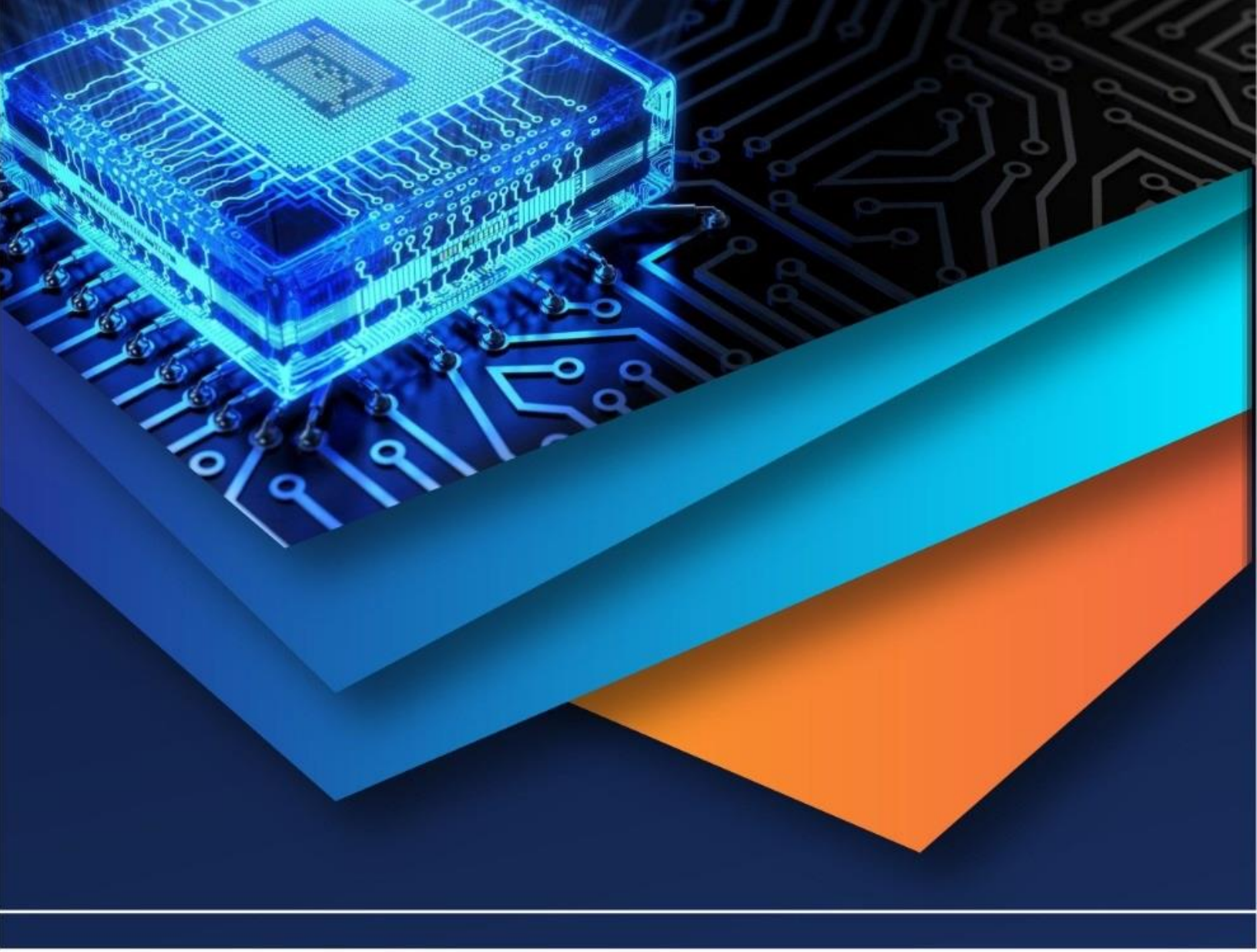
The graph further highlights the scalability and intelligent automation provided by the proposed framework, making it more suitable for modern cybercrime reporting and proactive threat prediction environments.

VII. CONCLUSION

This paper introduces an AI-powered system that automatically reports and predicts cybercrimes. It combines various reporting methods, uses machine learning to classify crimes, and provides real-time information about possible threats. The system makes it easier and faster to report cybercrimes, helps respond quicker, and improves how users and police work together. The system is very accurate in identifying different types of crimes and predicting potential threats, showing how well it works with cybercrime data. Features like creating organized reports, checking evidence, and sending real-time alerts help make the system more reliable and useful in actual situations. In general, this system provides a scalable and efficient way to manage cybercrime. Future plans include developing a mobile app, using blockchain for verifying evidence, and connecting with national cybercrime databases.

REFERENCES

[1] N. Geetha et al., "Cyberspace News Prediction of Text and Image with Report Generation," in Proc. IEEE Int. Conf. Comput. Commun. Signal Process. (ICCSP), 2020.
[2] S. Li et al., "Cybercrime Analysis Based on Bayesian Model," in Proc. Int. Symp. Big Data Appl. Serv. (ISBDAS), 2025.
[3] G. S. et al., "Predicting Cyber-Attacks Using Machine Learning Techniques," in Proc. Int. Conf. Smart Technol. Syst. Next Gener. Comput. (ICSTSN), 2024.
[4] A. R. et al., "Cyber Crime Prediction using Random Forest," in Proc. Int. Conf. Signal Process. Commun. Syst. (CSITSS), 2024.
[5] Veena K. et al., "Cybercrime Identification Using Machine Learning Techniques," Computational Intelligence and Neuroscience, vol. 2022, pp. 1–10, 2022.
[6] M. Chowdhury et al., "Blueprint: A Cyber Crime and Police Assistance Application," in Proc. Int. Conf. Data Sci. Bus. Syst. (ICDSBS), 2025.
[7] A. Mahmoud et al., "Online Crime Reporting System for Digital Forensics," in Proc. Int. Conf. Quality in Comput., Commun. Health, Eng. Smart Syst. (iQ-CCHESS), 2023.
[8] G. Al-Rummana et al., "Big Data Analysis in Crime Prediction," Wiley, 2021.
[9] Z. Abbass et al., "Social Crime Prediction using Twitter," in Proc. Int. Conf. Smart Comput. (ICSC), 2020.
[10] S. Deshmukh and P. Kamble, "CNN-LSTM Based Cyber Threat Detection," in Proc. Int. Conf. Adv. Intell. Syst. (ICAIS), 2023.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY