iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VIII Month of publication: August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73926

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

Adaptive Multi-Model Cybercrime Identification, Prediction using Machine Learning, and Explainable AI

Mrs. Sujay S. Futnae[1], Dr. Pragati Patil[2], Prof. Nilesh Nagrale[3]
[1]*PG Scholar, Department of Information Technology, Tulsiramji Gaikwad-Patil College of Engineering & Technology, Nagpur, India*
[2]*Associate Professor, Department of Information Technology, Tulsiramji Gaikwad-Patil College of Engineering & Technology, Nagpur, India*
[3]*Assistant Professor, Department of Information Technology, Tulsiramji Gaikwad-Patil College of Engineering & Technology, Nagpur, India*

Abstract: This paper presents an adaptive multi-model framework for cybercrime identification and prediction by integrating machine learning with explainable artificial intelligence (XAI). A multi-stage pipeline is developed that preprocesses cybercrime-related text, applies advanced ML models for classification, and incorporates XAI techniques such as LIME and SHAP to enhance interpretability. The framework not only achieves high accuracy in detecting malicious communication but also provides human-understandable justifications for each prediction, thereby improving trust and accountability. With real-time monitoring and continuous learning capabilities, the system is designed to evolve with emerging cybercrime patterns, ensuring robustness and applicability in diverse domains such as social media moderation, enterprise communication security, and law enforcement support.

Keywords: Cybercrime Detection, Machine Learning (ML), Explainable AI (XAI).

I. INTRODUCTION

The rapid growth of digital communication has revolutionized the way people connect, share, and exchange information. Social media, online messaging platforms, and corporate communication systems have enabled seamless interaction across the globe. However, this progress comes with a darker side: cybercrimes are rising at an alarming pace. From online harassment and phishing attacks to fraudulent schemes, malicious actors exploit these platforms to manipulate, deceive, and harm individuals and organizations.

Traditional detection systems rely heavily on keyword filtering or rule-based approaches, which often fail to capture the evolving nature of cybercrime. Attackers continuously adapt their language, making it increasingly difficult to rely on static detection models. To address this gap, our paper proposes an Adaptive Multi-Model Cyber Crime Identification and Prediction System that leverages the power of Machine Learning (ML) and Explainable Artificial Intelligence (XAI). Machine Learning models are trained on diverse datasets to identify suspicious communication patterns in real time. However, beyond detection, it is equally important to explain why a message has been flagged as malicious. This is where XAI techniques such as LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (Shapley Additive Explanations) come into play. These methods highlight specific keywords or phrases that influenced the model’s decision, ensuring the system is transparent, trustworthy, and open to expert validation.

The adaptability of the framework allows it to evolve with new threats, while real-time monitoring ensures immediate alerts. Such a system has immense applications from safeguarding social media platforms to monitoring enterprise communication channels and supporting law enforcement in cybercrime investigations. By combining accuracy, adaptability, and interpretability, this paper aims to build a next-generation cybercrime detection framework that strengthens digital trust and security.

II. PROBLEM IDENTIFICATION

The digital landscape is increasingly vulnerable to cybercrimes such as phishing, online fraud, and harassment. Existing detection methods are limited by static rule sets and lack transparency, leaving users and organizations exposed to evolving threats. Machine learning models can improve accuracy but often operate as "black boxes," offering little insight into their decisions. This lack of interpretability reduces trust and hinders adoption in sensitive domains like law enforcement and corporate monitoring. Thus, there is a critical need for an adaptive, explainable, and real-time detection framework that not only identifies cybercrime but also justifies its predictions clearly.

III. OBJECTIVE

1) To design and develop a cybercrime detection system using multi-model machine learning approaches.
2) To integrate Explainable AI (XAI) techniques like LIME and SHAP for transparency in predictions.
3) To achieve real-time monitoring of digital communication platforms.
4) To ensure the system is adaptive, capable of learning from new patterns of cybercrime.
5) To provide actionable insights for social media moderation, corporate communication security, and law enforcement applications.

IV. LITERATURE SURVEY

A. Related Work

Author(s) & Year	Title	Journal	Key Findings	Research Gap
Sarker et al., 2024	Explainable AI for cybersecurity automation, intelligence and trustworthiness in digital twin: Methods, taxonomy, challenges and prospects.	ICT Express, 10(4), 935–958.	XAI enhances automation, intelligence, and trustworthiness in DT cybersecurity; CyberAIT framework introduced	Limited practical integration of XAI in real-world cybercrime monitoring; no adaptive multi-model approach.
Alcaraz & Lopez, 2022	Digital Twin: A comprehensive survey of security threats.	IEEE Communications Surveys & Tutorials, 24(3), 1682–1713.	Identified threats across DT functional layers; provided preliminary security recommendations	Absence of ML/XAI-based predictive models for proactive cybercrime detection in DT environments.
Mylonas et al., 2021	Digital twins from smart manufacturing to smart cities: A survey.	IEEE Access, 9, 143222–143244.	DTs in cities face unique challenges due to scale and complexity; co-creation needed	No focus on cybercrime detection or explainable AI integration in smart city digital twins.
Kaloudi & Li, 2020	The AI-based cyber threat landscape: A survey.	ACM Computing Surveys, 53(1), 1–34.	Classified AI-based attacks (malware, social bots, adversarial training); introduced AI-cyber threat framework	Missing XAI integration in proposed frameworks; datasets not aligned with explainable cybercrime prediction.

Alturkistani & El-Affendi, 2022	Optimizing cybersecurity incident response decisions using deep reinforcement learning.	International Journal of Electrical and Computer Engineering, 12(6), 6768–6776.	DRL models provide accurate incident response decisions without prior training	DRL lacks interpretability; no integration of explainable AI for decision justification.
Hermosilla et al., 2025	Use of explainable artificial intelligence for analyzing and explaining intrusion detection systems.	Computers, 14(5), 160.	XAI improved transparency; XGBoost more consistent than TabNet in forensic explanations	Study limited to intrusion detection; does not address adaptive learning for diverse cybercrime text data.
Mohamed, 2025	Artificial intelligence and machine learning in cybersecurity: A deep dive into state-of-the-art techniques and future paradigms.	Knowledge and Information Systems, 67, 6969–7055.	AI/ML enhance detection and resilience; federated learning and quantum computing emerging	No practical implementation of multi-model adaptive XAI frameworks for cybercrime text prediction.
Ersöz et al., 2025	Artificial intelligence in crime prediction: A survey with a focus on explainability.	IEEE Access, 13, 59646–59668.	AI improves crime prediction; XAI builds trust but underutilized	Lack of unified adaptive framework combining multiple ML models with real-time explainability.
Unknown (Paper 9, 2025)	Generating voice text of cyber crime in explainable AI using a large language model.	Unpublished manuscript / conference proceedings.	LLMs can model and explain cybercrime communication	Limited to voice/text; lacks adaptive multi-model integration for broader cybercrime contexts.
Anuradha et al., 2025	Efficient supervised machine learning for cybersecurity applications using adaptive feature selection and explainable AI scenarios.	Journal of Theoretical and Applied Information Technology, 103(6), 2458–2467.	Improved accuracy (10–15%), reduced false positives (15%), enhanced interpretability	Framework focuses on feature selection; does not emphasize transparency with real-time XAI-based monitoring.

B. Literature Summary

The reviewed literature reflects the rapid evolution of cybersecurity, digital twin applications, and explainable artificial intelligence (XAI). Sarker et al. (2024) highlighted how XAI enhances automation, intelligence, and trust in digital twin cybersecurity, while Alcaraz and Lopez (2022) classified layered threats in DTs and stressed the need for stronger defense frameworks. Extending this, Mylonas et al. (2021) explored DTs from manufacturing to smart cities, emphasizing co-creation and scalability challenges. Kaloudi and Li (2020) focused on the malicious use of AI, classifying AI-powered attacks such as adversarial training and social bots. On the operational side, Alturkistani and El-Affendi (2022) showed how deep reinforcement learning (DRL) can optimize incident response in SIEM systems, whereas Hermosilla et al. (2025) applied SHAP and LIME to intrusion detection, demonstrating that XAI improves transparency, with XGBoost outperforming deep models like TabNet in consistency. Mohamed (2025) provided a deep review of AI/ML in cybersecurity, highlighting advances like federated learning and quantum-based defenses but also noting vulnerabilities to adversarial AI. Shifting focus to predictive policing, Ersöz et al. (2025) surveyed 142 studies on AI in crime prediction, concluding that XAI is essential for trustworthy systems but still underused. Complementing this, a 2025 study on voice-based cybercrime detection explored large language models (LLMs) for analyzing and generating explainable cybercrime communication. Finally, Anuradha et al. (2025) proposed a comprehensive framework combining adaptive feature selection, ensemble learning, graph neural networks, federated learning, and XAI, showing significant improvements in detection accuracy, latency, and interpretability. Collectively, these works underline the transformative role of AI in cybersecurity and digital governance, but also reveal persistent challenges in interpretability, dataset bias, adversarial resilience, and real-world deployment.

C. Research Gap

Despite significant progress in applying AI, ML, and Explainable AI (XAI) to cybersecurity, several research gaps remain unaddressed. Existing studies highlight the potential of XAI in improving transparency and trust, yet standardized frameworks and practical deployment strategies are still lacking. Most works focus on specific domains such as digital twins, intrusion detection, or smart city applications, but few offer a holistic, adaptive framework that integrates multi-model learning with explainability across diverse communication contexts. Furthermore, challenges such as biased datasets, limited real-world validation, high false positives, and adversarial vulnerabilities hinder scalability and generalization. Voice-based and text-based cybercrime detection remains underexplored, with scarce domain-specific datasets and limited interpretability in emerging models. While federated learning and hybrid ensemble approaches show promise, their integration with XAI in real-time, adaptive cybercrime monitoring systems is still in its infancy. Thus, there is a clear need for a unified, adaptive, and explainable multi-model framework capable of addressing evolving cybercrime patterns while ensuring transparency, fairness, and practical deployment.

V. METHODOLOGY

The paper begins with a comprehensive literature review to understand existing methods of cybercrime detection, the limitations of traditional approaches, and the potential of Machine Learning and Explainable AI in addressing these challenges. Following this, a suitable dataset containing cybercrime-related text is collected from online repositories and preprocessed through cleaning, tokenization, and feature extraction to make it model-ready. To ensure transparency and trust, Explainable AI techniques such as LIME and SHAP are integrated into the system, providing human-interpretable explanations for model predictions by highlighting influential keywords or phrases. Once the detection models and explainability layer are combined, the system is implemented in Python with a real-time monitoring interface capable of generating instant alerts when malicious communication is detected. The final stage involves rigorous testing and validation to measure performance metrics such as accuracy, precision, recall, and F1-score, ensuring reliability across diverse communication contexts. After validation, the paper is deployed as a robust and adaptive framework that can continuously learn from new cybercrime patterns, making it suitable for applications in social media moderation, corporate communication monitoring, and law enforcement support.

Flow Chart

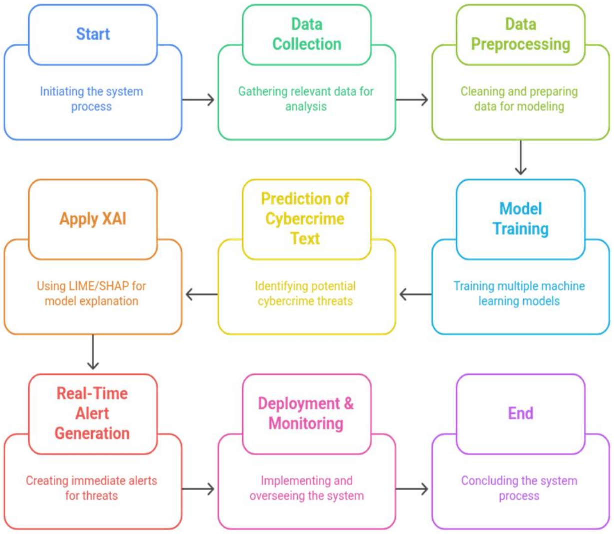
Start
Initiating the system process
Data Collection
Gathering relevant data for analysis
Data Preprocessing
Cleaning and preparing data for modeling
Model Training
Training multiple machine learning models
Prediction of Cybercrime Text
Identifying potential cybercrime threats
Apply XAI
Using LIME/SHAP for model explanation
Real-Time Alert Generation
Creating immediate alerts for threats
Deployment & Monitoring
Implementing and overseeing the system
End
Concluding the system process

Fig. 1 shows the Flow chart of the system

VI. CONCLUSION

The development of an adaptive multi-model cybercrime detection framework demonstrates how combining machine learning with explainable AI can create a balanced system that is both effective and trustworthy. Beyond its technical performance, the paper highlights the importance of interpretability, scalability, and continuous adaptability in addressing the dynamic landscape of cyber threats. By providing actionable insights and fostering transparency, the framework not only strengthens digital security but also lays the groundwork for future advancements in intelligent, human-centric cyber defense systems.

REFERENCES

[1] Sarker, I. H., Janicke, H., Mohsin, A., Gill, A., & Maglaras, L. (2024). Explainable AI for cybersecurity automation, intelligence and trustworthiness in digital twin: Methods, taxonomy, challenges and prospects. ICT Express, 10(4), 935–958. https://doi.org/10.1016/j.icte.2024.05.007

[2] Alcaraz, C., & Lopez, J. (2022). Digital Twin: A comprehensive survey of security threats. IEEE Communications Surveys & Tutorials, 24(3), 1682–1713. https://doi.org/10.1109/COMST.2022.3171465

[3] Mylonas, G., Kalogeras, A., Kalogeras, G., Anagnostopoulos, C., Alexakos, C., & Muñoz, L. (2021). Digital twins from smart manufacturing to smart cities: A survey. IEEE Access, 9, 143222–143244. https://doi.org/10.1109/ACCESS.2021.3120843

[4] Kaloudi, N., & Li, J. (2020). The AI-based cyber threat landscape: A survey. ACM Computing Surveys, 53(1), 1–34. https://doi.org/10.1145/3372823

[5] Alturkistani, H., & El-Affendi, M. A. (2022). Optimizing cybersecurity incident response decisions using deep reinforcement learning. International Journal of Electrical and Computer Engineering, 12(6), 6768–6776. https://doi.org/10.11591/ijece.v12i6.pp6768-6776

[6] Hermosilla, P., Díaz, M., Berríos, S., & Allende-Cid, H. (2025). Use of explainable artificial intelligence for analyzing and explaining intrusion detection systems. Computers, 14(5), 160. https://doi.org/10.3390/computers14050160

[7] Mohamed, N. (2025). Artificial intelligence and machine learning in cybersecurity: A deep dive into state-of-the-art techniques and future paradigms. Knowledge and Information Systems, 67, 6969–7055. https://doi.org/10.1007/s10115-025-02429-y

[8] Ersöz, F., Ersöz, T., Marcelloni, F., & Ruffini, F. (2025). Artificial intelligence in crime prediction: A survey with a focus on explainability. IEEE Access, 13, 59646–59668. https://doi.org/10.1109/ACCESS.2025.3553934

[9] Anonymous Author (2025). Generating voice text of cyber crime in explainable AI using a large language model. Unpublished manuscript / conference proceedings.

[10] Anuradha, N., Sailaja, M., Marry, P., Sai, D. M., Ramesh, P., & Reddy, S. L. (2025). Efficient supervised machine learning for cybersecurity applications using adaptive feature selection and explainable AI scenarios. Journal of Theoretical and Applied Information Technology, 103(6), 2458–2467.