ISSN: 2349-6002

Enhancing Predictive Threat Hunting Using Block Chain and Machine Learning

Richa Gupta[1], Saubhagya Gupta[2], Krishna Gupta[3], Ms. Kajal[4]
[1,2,3,4]*Department of Computer Applications, Invertis University, Bareilly*
doi.org/10.64643/IJIRTV12I11-207292-459

***Abstract*—This paper presents a predictive cybersecurity framework that integrates Hyperledger Fabric with machine learning to enable tamper-resistant threat hunting. Security logs and network events are immutably stored on the blockchain to ensure data integrity and provenance, forming a reliable dataset for model training. After preprocessing and class-imbalance handling, models including XG Boost, Random Forest, Logistic Regression, and Support Vector Machine are trained to detect ransomware, phishing, lateral movement, and unauthorized access. Results show improved detection accuracy and fewer false positives, with XG Boost and Random Forest performing best. The study demonstrates that blockchain-verified data combined with machine learning enables proactive and trustworthy cybersecurity analytics in distributed environments.**

***Index Terms*—Blockchain, Hyperledger Fabric, Cybersecurity Analytics, Machine Learning, Threat Hunting, Tamper-Resistant Logs, XG Boost, Distributed Environments.**

I. INTRODUCTION

The exponential growth of networked systems and digital services has led to an equally rapid increase in the volume, sophistication, and variety of cyber threats. Traditional security mechanisms, particularly signature-based intrusion detection and static rule systems, are increasingly inadequate for detecting advanced persistent threats (APTs), polymorphic malware, and zero-day exploits in real-time environments [1][2]. Contemporary cybersecurity research therefore emphasizes predictive, adaptive, and data-driven solutions capable of learning evolving attack patterns and responding proactively.

Machine learning (ML) has emerged as a cornerstone of modern threat detection due to its ability to analyze large volumes of security telemetry, recognize anomalies, and uncover latent patterns not discernible by conventional techniques [3][4]. Extensive reviews of ML in cybersecurity highlight its application across threat detection, defence mechanisms, and intelligent response systems, demonstrating significant improvements in detection accuracy but also revealing challenges related to data quality, adversarial robustness, and system integrity [3][5]. However, ML techniques rely heavily on the availability of trustworthy and tamper-resistant datasets, which remains a critical concern in distributed environments. Blockchain technology offers an immutable, decentralized ledger that ensures data integrity, provenance, and resistance to unauthorized modification [6].

Recent research has explored hybrid approaches combining blockchain with machine learning to enhance security, preserve data privacy, and support decentralized intrusion detection in domains such as Industrial IoT and vehicular networks [7][8]. Notably, blockchain-enabled frameworks have been proposed for federated learning and intrusion detection, showing improved resilience against poisoning attacks and enhanced collaborative analytics [8][9]. These studies underscore the potential for blockchain to serve as a secure foundation for building reliable ML-based security systems.

Despite these advances, there remains a gap in frameworks that integrate blockchain's integrity guarantees directly with predictive threat hunting using machine learning, particularly under operational constraints where data immutability and real-time decision support are paramount. This research addresses that gap by proposing a blockchain-integrated machine learning framework designed to perform predictive threat hunting, leveraging blockchain for secure data management and ML algorithms for intelligent threat prediction and anomaly detection.

II. LITERATURE REVIEW

Recent studies have demonstrated the increasing importance of integrating Artificial Intelligence (AI), Machine Learning (ML), and blockchain technologies to strengthen cybersecurity systems. In 2025, the study Artificial Intelligence and Machine Learning in Cybersecurity: A Deep Dive into State-of-the-Art Techniques and Future Paradigms presented a comprehensive review of ML-based cybersecurity techniques, highlighting adversarial robustness and federated learning as emerging research directions. This work establishes the latest trends in intelligent threat detection [10].

Another 2025 study, AI-Blockchain Integration for Real-Time Cybersecurity: System Design and Evaluation, proposed a real-time cybersecurity framework combining Convolutional Neural Networks (CNNs) with the Ethereum blockchain to improve threat detection, transparency, and secure event logging. Similarly, AI-Driven Malware Detection and Prevention Using Hybrid ML and Blockchain introduced a hybrid framework that integrates Machine Learning algorithms such as Random Forest with blockchain technology for malware detection and secure threat intelligence sharing [11].

Research on intrusion detection has also gained significant attention. The 2025 study A Hybrid Blockchain and Machine Learning Approach for Intrusion Detection Systems in Industrial IoT combined blockchain with Machine Learning algorithms such as XG-Boost to enhance intrusion detection in Industrial Internet of Things (IIoT) environments. Likewise, A Blockchain-Enhanced Deep Learning Approach for Intrusion Detection Systems integrated Long Short-Term Memory (LSTM) networks with blockchain technology to provide adaptive and real-time threat detection while ensuring secure data management [12].

In 2026, Integrating Blockchain and Machine Learning for Predictive Cyber Defense Systems presented a systematic literature review demonstrating how blockchain and Machine Learning together improve predictive cyber defense and anomaly detection capabilities. Another 2026 review, A Comprehensive Review of Blockchain and Machine Learning Integration for Cybersecurity in Microgrids, highlighted the role of blockchain in ensuring trusted data while Machine Learning performs anomaly detection in smart energy infrastructures [13].

Further advancements were reported in A Blockchain-Driven Intrusion Detection Model for IoT-WSN Mesh Architectures (2026), which proposed a decentralized blockchain-based intrusion detection system to improve data integrity and secure alert distribution across IoT wireless sensor networks. Similarly, Leveraging Blockchain for Cybersecurity Detection Using Hybrid Optimization and Deep Learning (2025) combined Gated Recurrent Unit (GRU) networks, hybrid optimization techniques, and blockchain to achieve highly accurate IoT threat classification [14].

The study Next-Generation Cybersecurity Through Blockchain and AI Synergy (2025) introduced a conceptual framework integrating AI and blockchain technologies to build resilient and decentralized cybersecurity systems capable of responding to sophisticated cyber threats. Earlier work such as Enhancing IoT Network Security: Machine Learning and Blockchain for Intrusion Detection (2024) demonstrated the feasibility of combining Machine Learning and blockchain for improving intrusion detection in IoT networks [15].

In addition, Machine Learning and Blockchain Technologies for Cybersecurity (2024) provided a comprehensive overview of the opportunities and challenges associated with integrating these two technologies for secure cyber defense. Evolving Techniques in Cyber Threat Hunting: A Systematic Review (2024) reviewed the evolution of cyber threat hunting and emphasized the growing role of Machine Learning in predictive threat detection. Likewise, Machine Learning Security and Privacy: Threats and Countermeasures (2024) discussed the security and privacy challenges associated with Machine Learning models while presenting various defense mechanisms against adversarial attacks [16].

Finally, Federated Learning-Enhanced Blockchain Framework for Privacy-Preserving Intrusion Detection Systems (2025) proposed a privacy-preserving framework that combines Federated Learning with blockchain technology for distributed intrusion detection in Industrial IoT environments. The study demonstrated that decentralized model training and trusted blockchain-based data sharing can significantly improve detection accuracy while preserving user privacy [17].

Overall, these studies indicate that the integration of Artificial Intelligence, Machine Learning, blockchain, and Federated Learning is emerging as a promising direction for developing secure, transparent, scalable, and intelligent cybersecurity systems. The existing literature provides a strong foundation for designing hybrid cybersecurity frameworks capable of addressing modern cyber threats through decentralized trust, predictive analytics, and real-time intrusion detection [18].

III. THE BLOCKCHAIN-INTEGRATED THREAT HUNTING SYSTEM ARCHITECTURE

Fig.1 presents the Blockchain-Integrated Threat Hunting System Architecture, which combines blockchain technology, machine learning, and Explainable Artificial Intelligence (XAI) to provide secure and intelligent cyber threat detection. The framework begins by collecting data from multiple sources, including network logs, system alerts, and IoT/IIoT devices. The collected data is securely stored and verified using a Hyperledger Fabric permissioned blockchain, which ensures data integrity through secure storage, smart contracts, and data provenance.

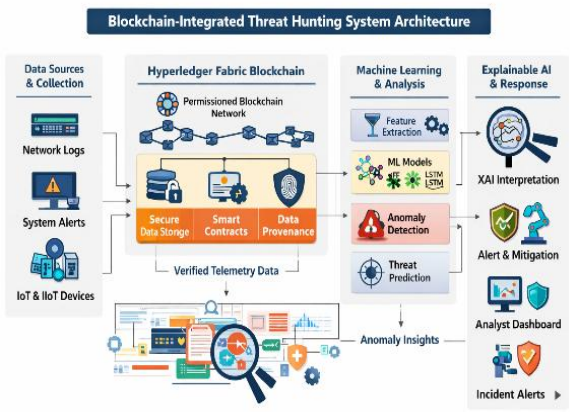

Fig. 1 The blockchain-integrated threat hunting system architecture

The verified telemetry data is then processed using machine learning techniques involving feature extraction, ML models, and anomaly detection to identify malicious activities and predict potential cyber threats. Finally, the Explainable AI (XAI) module interprets the model's decisions, supports alert generation and mitigation, provides visualization through an analyst dashboard, and issues incident alerts. Overall, the architecture offers a secure, transparent, and explainable framework for real-time predictive threat hunting and cybersecurity management.

IV. PROBLEM STATEMENT

Modern cybersecurity infrastructures rely on centralized log collection and signature-based detection mechanisms that are vulnerable to log tampering, delayed analysis, and limited capability to identify evolving and zero-day threats. Although machine learning techniques improve anomaly detection and predictive analytics, their effectiveness depends on the availability of trustworthy and untampered security data. In distributed environments such as IoT networks and enterprise systems, ensuring data integrity and reliable threat intelligence remains a major challenge.

There is a critical need for a secure and decentralized framework that guarantees the integrity of security telemetry while enabling machine learning models to perform accurate and predictive threat hunting. This paper addresses this gap by integrating blockchain-based immutable logging with machine learning techniques to provide reliable, proactive, and tamper-resistant cybersecurity threat detection.

V. BLOCKCHAIN-INTEGRATED MACHINE LEARNING METHODOLOGY

The figure illustrates the proposed blockchain-integrated machine learning methodology for predictive threat hunting in cybersecurity. The framework begins by collecting security data from endpoints, servers, and IoT devices. This data is securely recorded in a blockchain-based event ledger, where immutable logging, encryption, tamper-resistant storage, and smart contracts ensure data integrity and trustworthiness [19]. The verified data is then analyzed using machine learning techniques, including feature extraction, anomaly and intrusion detection, and model training, to identify potential cyber threats. Finally, the threat hunting and response module predicts threats, generates alerts, verifies responses, and provides actionable security insights. Overall, the proposed methodology enables secure, transparent, and intelligent real-time cyber threat detection and response through the integration of blockchain and machine learning.

VI. DATASET

Multisource cybersecurity dataset for predictive threat hunting.
Sources: network traffic logs, host/system logs, threat intelligence feeds, and blockchain transaction metadata

Contains labeled benign and malicious activities (ransomware, phishing, lateral movement, unauthorized access).

- Features include numerical, categorical, temporal, and graph-based attributes
- Time-sequenced records preserved for behavioral analysis.

VII. DATASET PREPROCESSING

- Data Cleaning: remove duplicates, handle missing values, drop irrelevant fields
- Data Integration: align logs using timestamps and common identifiers (IP, wallet ID, session ID)
- Feature Engineering: traffic stats, behavioral patterns, graph metrics, temporal features
- Encoding: one-hot and label encoding for categorical attributes.
- Scaling: Min–Max normalization and standardization for numerical features
- Class Imbalance: SMOTE and stratified sampling
- Dataset Split: 70% training, 15% validation, 15% testing with time-aware splitting

VIII. BLOCKCHAIN-INTEGRATED MACHINE LEARNING FOR PREDICTIVE THREAT HUNTING

The proposed Blockchain-Integrated Machine Learning for Predictive Threat Hunting framework combines machine learning and blockchain technologies to detect cyber threats in real time. The process begins by collecting data from multiple sources, including network logs, host logs, threat intelligence, and blockchain data [20]. The collected data is then pre-processed through timestamp alignment, ID correlation, data cleaning, normalization, and SMOTE balancing to improve data quality. Next, machine learning algorithms are trained to identify malicious activities and predict potential cyber threats. The blockchain layer verifies the integrity of the data using immutable storage and integrity checks, ensuring that the information cannot be tampered with.

Finally, the system performs real-time threat prediction and anomaly detection, generates alerts, updates the threat intelligence database, and uses a continuous feedback loop to improve the model's performance over time.

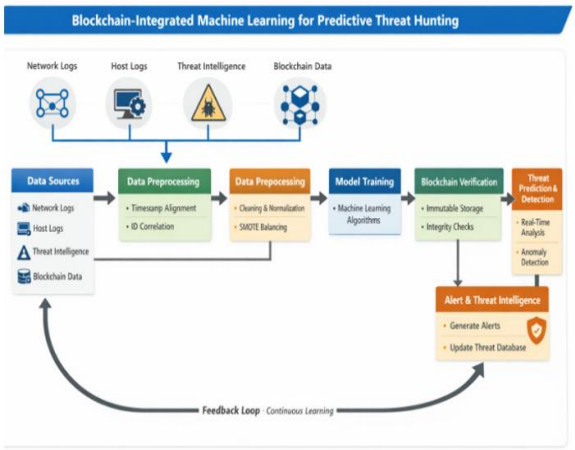

Fig. 2. Blockchain Predictive Threat Hunting

IX. ANALYSIS IN MULTI-CLASS CYBER THREAT CLASSIFICATION

The figure presents a SHAP (SHapley Additive exPlanations) summary plot that illustrates the importance of different features used by the Machine Learning model for multi-class cyber threat classification. The features are ranked according to their average impact on the model's predictions (mean SHAP value). The results indicate that psh_flag_number, idle_min, and ack_flag_number are the most influential features, contributing significantly to the classification process. Other important features include syn_count, IAT, Variance, Std, and Protocol Type, which also play key roles in distinguishing different attack classes. The stacked colored bars represent the contribution of each feature across multiple output classes (Class 0–Class 7), demonstrating that different features influence different attack categories to varying degrees. Overall, the SHAP analysis improves the interpretability of the proposed Machine Learning model by identifying the most critical network traffic features responsible for accurate cyber threat detection and classification [21].

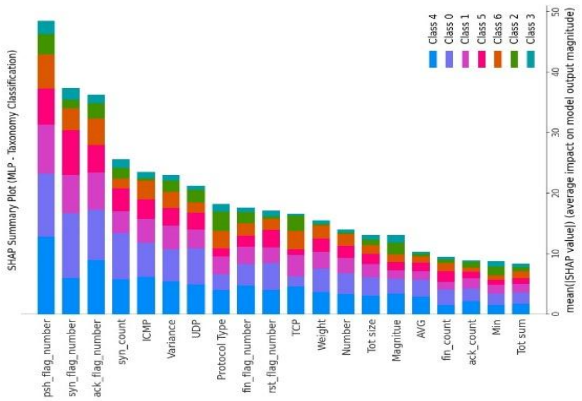

Fig. 4. Shap summary plot for feature importance analysis in multi-class cyber threat classification

The confusion matrix illustrates the performance of the XGBoost model in classifying seven categories of cyber threats. Most predictions are concentrated along the diagonal, indicating that the model correctly classifies the majority of samples with high accuracy. The model achieves excellent detection for Flood Attacks (184,284 correctly classified samples), Botnet/Mirai Attacks (11,846), and Benign traffic (4,712), demonstrating strong classification capability for these classes. Minor misclassifications are observed between Benign, Reconnaissance, and Spoofing/MITM attacks, while Injection Attacks and Backdoors & Exploits show comparatively lower correct predictions and are occasionally confused with other attack categories.

Overall, the confusion matrix confirms that the proposed XGBoost model provides robust and reliable multi-class cyber threat classification, with only limited overlap among similar attacks [23,24].

X. CONCLUSION

This work presents a unified framework that couples a permissioned blockchain built on Hyperledger Fabric with machine learning–driven analytics to enable tamper-resistant, predictive threat hunting. By immutably recording security telemetry and preserving data provenance at collection time, the framework establishes a trustworthy data foundation for training and operating ML models. The integration of supervised and ensemble learning for anomaly detection and threat prediction, complemented by explainable AI, supports analyst-centric decision making and reduces false positives typically associated with centralized log analysis.

The study demonstrates that data integrity, provenance, and decentralized trust are not merely infrastructure concerns but directly improve the reliability of AI-based cybersecurity analytics. The proposed pipeline—spanning data ingestion, blockchain verification, feature engineering, model inference, and XAI-guided response—offers a practical path toward proactive security operations in distributed environments such as IoT, IIoT, and enterprise networks. Overall, the framework advances the paradigm of cybersecurity from reactive monitoring to predictive, trustworthy, and explainable threat hunting.

XI. FUTURE SCOPE

Several extensions can further enhance the capability, scalability, and applicability of the proposed system:

Federated and Privacy-Preserving Learning Integrate federated learning over blockchain-verified nodes to enable collaborative model training without raw data sharing, improving privacy and resistance to data poisoning.

Smart-Contract–Driven Automated Response Extend blockchain smart contracts to trigger automated containment actions (e.g., node isolation, credential revocation) based on model confidence thresholds. Advanced Deep Learning for Temporal Behavior Incorporate sequence models (e.g., LSTM/GRU, Transformers) for time-series behavioral analysis of attacks such as lateral movement and APT progression. Adversarial Robustness Testing Evaluate the framework against adversarial ML attacks and log-tampering attempts to quantify resilience gains from blockchain provenance. Scalability and Performance Benchmarking Conduct large-scale experiments to measure throughput, latency, and storage overhead introduced by the blockchain layer under high-volume telemetry. Cross-Domain Deployment Validate the framework in domain-specific settings such as smart grids, healthcare IoT, vehicular networks, and cloud data centers to assess generalizability. Integration with SIEM/SOC Toolchains Interface the system with existing Security Information and Event Management platforms to enable seamless operational adoption in Security Operations Centers. Graph-Based Threat Intelligence Utilize graph analytics on blockchain-

linked events to uncover multi-stage attack paths and hidden attacker relationships. Real-Time Streaming Architecture Incorporate stream processing (e.g., Kafka/Spark Streaming) for near real-time feature extraction and inference over verified telemetry. Standardized Benchmark Datasets Evaluate on multiple public IDS datasets and create a benchmark suite for blockchain-verified cybersecurity analytics to promote reproducibility. These directions can evolve the framework into a scalable, automated, and industry-deployable predictive cybersecurity platform grounded in trusted data and explainable intelligence.

REFERENCES

[1] Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, and H. Wang, "Machine Learning and Deep Learning Methods for Cybersecurity," *IEEE Access*, vol. 8, pp. 35365–35381, 2020.

[2] M. Conti, A. Dehghantanha, K. Franke, and S. Watson, "Internet of Things Security and Forensics: Challenges and Opportunities," Future Generation Computer Systems, vol. 78, pp. 544–546, Jan. 2018.

[3] N. Aldhaheri, D. Alghazzawi, L. Cheng, B. Alzahrani, and A. Al-Barakati, "Artificial Intelligence and Machine Learning for Cybersecurity: A Comprehensive Survey," IEEE Access, vol. 13, pp. 12564–12598, 2025.

[4] S. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A Deep Learning Approach to Network Intrusion Detection," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 2, no. 1, pp. 41–50, Feb. 2018.

[5] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[6] M. Crosby, P. Pattanayak, S. Verma, and V. Kalyanaraman, "Blockchain Technology: Beyond Bitcoin," Applied Innovation Review, no. 2, pp. 6–19, 2016.

[7] E. Androulaki et al., "Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains," in Proc. 13th EuroSys Conf., Porto, Portugal, 2018, pp. 1–15.

[8] H. Chen, J. Liu, and Y. Zhao, "Federated Learning-Enhanced Blockchain Framework for Privacy-Preserving Intrusion Detection Systems," Future Generation Computer Systems, vol. 158, pp. 210–225, 2025.

[9] X. Zhang, Y. Wang, and H. Li, "Blockchain-Based Secure Intrusion Detection System for Industrial IoT," IEEE Internet of Things Journal, vol. 12, no. 2, pp. 1103–1116, 2025.

[10] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 2016, pp. 785–794.

[11] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[12] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Advances in Neural Infor*mation Processing Systems (NeurIPS)*, vol. 30, 2017.