

## Cyber Threat Intelligence System

Shekhar Mishra<sup>1</sup>, Surya Pratap Singh<sup>2</sup>, Suryam Giri<sup>3</sup>

<sup>1,2,3</sup>Department of Computer Science & Engineering, BBDITM, Lucknow

**EmailID** shekharmishra1221@gmail.com<sup>1</sup>, suryaprataps471@gmail.com<sup>2</sup>,

Asuryamgoswami619@gmail.com<sup>3</sup>,

### Abstract

*In recent years, the rapid growth of digital technologies has led to an increase in sophisticated cyber threats, making traditional signature-based detection systems insufficient for modern cybersecurity needs. This paper presents the Cyber Threat Intelligence Dashboard, an intelligent and real-time cyber threat detection system that integrates machine learning, feature-based analysis, and threat intelligence into a unified platform. The system focuses on analyzing Indicators of Compromise (IOCs), such as IP addresses and domain names, by extracting meaningful structural and statistical features including length, character distribution, and entropy. The extracted features are utilized to train multiple machine learning models, such as Random Forest, Decision Tree, and Logistic Regression, for accurate classification of indicators into safe and malicious categories. A comparative analysis of these models is performed using evaluation metrics such as accuracy, precision, recall, and F<sup>1</sup>-score to identify the most effective approach. Overall, the Cyber Threat Intelligence Dashboard offers a scalable, efficient, and intelligent approach to cyber threat detection, with the potential for future enhancements in predictive analytics and large-scale deployment.*

**Keywords:** Cyber Threat Intelligence, Machine Learning, Feature Extraction, Real-Time Detection, Indicator of Compromise, Cybersecurity Dashboard

### 1. Introduction

In the digital era, the rapid growth of internet usage has significantly increased exposure to cyber threats such as phishing attacks, malicious domains, and suspicious IP activities. Traditional security mechanisms, which rely heavily on signature-based detection, often fail to identify newly emerging or obfuscated threats. This limitation highlights the need for intelligent and adaptive systems capable of analyzing and classifying potential threats in real time. The Cyber Threat Intelligence System is designed to address this challenge by integrating machine learning with real-time threat intelligence. The system focuses on analyzing Indicators of Compromise (IOCs), such as IP addresses and domain names, and classifying them as either benign (safe) or malicious. By leveraging feature extraction techniques—including structural attributes (length, number of digits, special characters), domain-based characteristics (top-level domain, subdomain length), and statistical measures like entropy—the system transforms raw indicators into meaningful numerical representations suitable for machine learning models.

To enhance detection accuracy, multiple classification algorithms are evaluated, including Random Forest, Decision Tree, and Logistic Regression. These models are trained on curated datasets and compared based on performance metrics such as accuracy, precision, and recall. The best-performing model is then deployed within a backend system to provide real-time predictions. In addition to machine learning capabilities, the project incorporates real-time threat data obtained from external intelligence platforms, enabling dynamic monitoring and visualization of cybersecurity risks. The frontend dashboard presents this information through interactive visualizations, allowing users to gain insights into threat patterns and model performance. Overall, this project aims to bridge the gap between static cybersecurity tools and intelligent, data-driven threat detection systems by providing an integrated platform for analysis, prediction, and visualization of cyber threats.

#### 1.1. Challenges in Existing Ride Coordination Systems

Despite significant advancements in cybersecurity, existing threat detection systems face several

limitations that reduce their effectiveness in identifying modern and evolving cyber threats. These challenges highlight the need for more adaptive, intelligent, and data-driven approaches.

- **Reliance on Signature-Based Detection:** Traditional systems depend on predefined signatures of known threats, making them ineffective against new or unknown (zero-day) attacks.
- **Lack of Real-Time Detection:** Many existing solutions analyze data in batches, causing delays in identifying threats and increasing the risk of damage.
- **Inability to Detect Evolving Threats:** Cyber attackers continuously modify attack patterns, which makes it difficult for static systems to keep up with new threat variations.
- **Limited Feature-Based Analysis:** Conventional approaches do not deeply analyze structural and statistical features such as domain entropy, character patterns, or URL behavior.

## 1.2. Introduction of Cyber Threat Intelligence System

To address these challenges, the Cyber Threat Intelligence System is introduced as an intelligent cyber threat analysis platform that combines machine learning, feature-based detection, and real-time threat intelligence within a unified system. The platform focuses on analyzing indicators such as IP addresses and domain names by extracting meaningful structural and statistical features, enabling accurate classification of threats as safe or malicious. By integrating real-time data sources with a scalable backend and an interactive dashboard interface, the system ensures timely detection, improved decision-making, and enhanced visibility into emerging cyber threats. Emphasizing adaptability, accuracy, and ease of use, the proposed solution provides a modern approach to cybersecurity that overcomes the limitations of traditional detection mechanisms without requiring complex infrastructure.

## 2. Theoretical Framework

### 2.1. Indicator Of Compromise (IOC) Analysis

The system is based on the concept of analyzing

Indicators of Compromise (IOCs), such as IP addresses and domain names, to identify potential cyber threats. These indicators serve as the primary input for the system and represent real-world entities that may exhibit malicious behavior. Proper analysis of IOCs enables early detection of suspicious activities in network environments.

### 2.2. Feature Extraction And Representation

Raw indicators cannot be directly used for machine learning; therefore, the system applies feature extraction techniques to convert them into structured numerical data. Features such as length, number of digits, special characters, number of dots, top-level domain length, subdomain length, and entropy are derived. These features capture both structural and statistical properties, enabling better differentiation between safe and malicious indicators.

### 2.3. Spatial Distance Modeling And Safety Thresholds

The extracted features are used to train supervised machine learning models for classification. Algorithms such as Random Forest, Decision Tree, and Logistic Regression are employed to learn patterns from labeled data. These models classify indicators into categories such as safe or malicious based on learned relationships between features and outcomes.

### 2.4. Model Evaluation And Performance Analysis

To ensure the effectiveness of the system, multiple models are evaluated using performance metrics such as accuracy, precision, recall, and F1-score. Comparative analysis helps in selecting the best-performing model for deployment. Additionally, feature importance analysis provides insights into which features contribute most to the prediction process.

### 2.5. Real-Time Integration And System Deployment

The final component of the framework involves integrating the trained model into a real-time system using a backend API. The system processes incoming indicators, performs predictions instantly, and delivers results to a frontend dashboard. Integration with external threat intelligence sources further enhances the system by providing up-to-date information on emerging cyber threats.

### 3. Literature Review

Ref. No	Author(s) & Year	Focus Area	Key Contribution	Identified Limitations
[1]	Santos, P., et al. (2025)	Cyber Threat Intelligence (CTI) Review	Provides a comprehensive systematic review of various CTI approaches, highlighting data sources, processing techniques, and intelligence sharing mechanisms.	Primarily focuses on theoretical analysis and lacks implementation-level validation.
[11]	Sharma, R., & Lee, J. (2024)	Deep Learning for Threat Profiling	Proposes deep learning-based models for threat profiling and classification, demonstrating improved accuracy in identifying complex cyber threats using neural networks.	High computational complexity and lack of explainability.

[14]	Park, T., et al. (2024)	Graph Neural Networks	Models relationships between cyber threats for better correlation.	High complexity and poor real-time performance.
------	-------------------------	-----------------------	--	---

### 4. Research Gap

Although existing mobility and ride-sharing systems have improved navigation and convenience, significant research gaps remain before such platforms can effectively support safe, reliable, and coordinated group travel. Current solutions largely prioritize individual optimization and commercial efficiency, leaving critical aspects of group-level safety, reliability, and integration insufficiently addressed.

#### 4.1. Limited Use Of Feature-Based Detection

The most Several studies, such as Santos et al. (2025) [1] and Khan et al. (2025) [4], highlight that most existing cybersecurity systems still rely heavily on signature-based and rule-based detection techniques. While these methods are effective for identifying known threats, they lack the ability to generalize and detect newly emerging or obfuscated malicious indicators. Even though machine learning has been introduced in some works, the focus is often limited to basic features rather than deep structural and statistical characteristics such as entropy, character randomness, and domain complexity. As highlighted in studies like Chen et al. (2024) [21], feature engineering plays a crucial role in improving detection accuracy, yet it remains underutilized. This creates a significant gap in developing robust feature-driven models capable of adapting to evolving cyber threats.

#### 4.2. Lack Of Integration Between Machine Learning And Real-Time Intelligence

Research by Johnson et al. (2025) [6] and Fernandez et al. (2024) [17] emphasizes the growing importance of real-time threat intelligence using OSINT and automated threat-sharing platforms. However, most existing systems either focus on real-time data collection or on offline machine learning

models, with very limited integration between the two. Machine learning models trained on static datasets often fail to adapt to rapidly changing threat landscapes, while real-time systems lack predictive intelligence. Furthermore, studies like Kumar & Das (2024) [18] show that even advanced SIEM–SOAR systems struggle to fully integrate predictive analytics with live threat feeds. This lack of synergy presents a critical research gap in building systems that combine real-time intelligence with adaptive machine learning for proactive threat detection.

#### 4.3. Insufficient Model Comparison And Evaluation

Several research works, including Patel & Roy (2025) [5] and Li & Fernandez (2024) [9], demonstrate the application of machine learning techniques in cybersecurity. However, many of these studies rely on a single algorithm or provide limited comparative analysis between different models. Even when multiple models are considered, detailed evaluation using performance metrics such as precision, recall, F1-score, and confusion matrix is often missing. According to Wang et al. (2024) [12], comprehensive model evaluation is essential for ensuring robustness and generalizability. The absence of systematic benchmarking and comparison frameworks limits the ability to identify the most suitable model for specific threat detection tasks, highlighting a major gap in existing research.

#### 4.4. Absence Of Unified Platforms

Research by Singh et al. (2023) [34] and Kumar & Das (2024) [18] discusses various components of cybersecurity systems, including threat visualization dashboards, SIEM, and SOAR frameworks. However, these components are typically developed and deployed as separate modules rather than as part of an integrated system. This fragmentation increases system complexity, reduces usability, and requires significant manual effort to correlate data from different sources. Studies like Costa et al. (2024) [16] also highlight the challenges of data normalization across platforms. There is a clear research gap in developing unified platforms that seamlessly integrate data collection, feature extraction, machine learning prediction, and visualization within a single cohesive system.

#### 4.5. Limited Focus On Explainability

Recent advancements in AI-driven cybersecurity, as discussed by Zhou et al. (2024) [15] and Sharma & Lee (2024) [11], have significantly improved detection capabilities using deep learning and advanced models. However, these models often operate as black boxes, providing predictions without clear explanations. This lack of interpretability reduces trust among cybersecurity analysts and makes it difficult to validate or justify decisions. As highlighted in Zhao & Lin (2023) [37], explainability is becoming increasingly important for real-world adoption of AI systems. Despite this, many existing solutions do not incorporate feature importance analysis or explainable AI techniques, indicating a critical research gap.

### 5. Proposed System

The proposed Cyber Threat Intelligence System aims to build a unified, research-oriented platform for real-time threat monitoring, machine-learning-driven threat categorization, and interactive visual analytics. This project's objective is to close the gap between scattered threat sources and practical intelligence by bringing together data collection, preprocessing, analytical modeling, visualization, and user interaction into a single coherent framework. This system is meant to support both research and operational needs, enabling analysts, students, and cybersecurity professionals to work with live threat streams, compare machine learning models, and observe global threat patterns in a dynamic manner. Unlike enterprise-restricted solutions, this dashboard encourages open-source usability and adaptability for continuous enhancement.

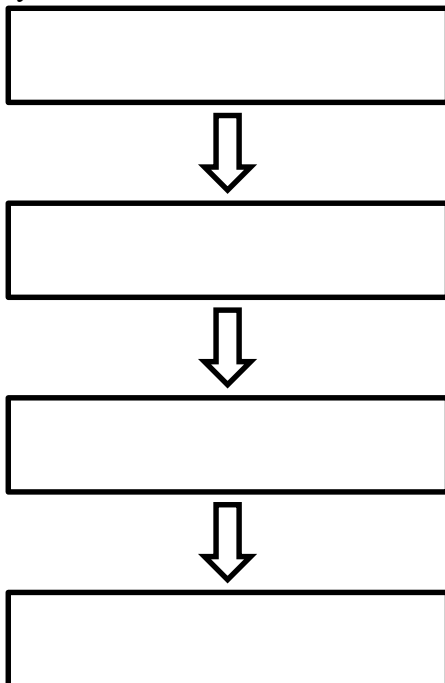
#### System Overview-

The entire system workflow is divided into five interconnected modules:

- Data Acquisition Layer
- Data Preprocessing Layer
- Machine Learning And Analysis Layer
- Visualization And Dashboard Layer
- User Interaction Layer

Together, these modules enable end-to-end processing from fetching live data to producing meaningful, interpretable intelligence for decision-making. The workflow ensures smooth data flow — from acquisition to visualization and maintaining

both speed and analytical depth. Together, these modules create a seamless pipeline that transforms raw, unstructured threat data into meaningful analytical insights and interactive visualizations. The entire workflow follows a logical progression — starting from data acquisition, moving through data preparation and model analysis, and finally reaching visualization and user interaction. This layered design ensures that data flows smoothly from one stage to another, maintaining both analytical precision and processing efficiency. The modular structure offers flexibility for future enhancements, allowing new data sources, advanced machine learning techniques, or visualization components to be integrated without disrupting the existing architecture. The system smoothly handles every stage, from data acquisition to visualization, keeping both speed and accuracy intact. It turns raw threat data into meaningful insights and interactive visuals through a logical, step-by-step pipeline. This layered structure ensures efficient data movement and precise analysis at every stage without affecting existing system's flow overall.



### 5.1. System Architecture

The proposed Cyber Threat Intelligence Dashboard follows a modular and layered architecture designed to efficiently handle data collection, processing, prediction, and visualization. The system integrates

machine learning with real-time threat intelligence to provide accurate and scalable cyber threat detection.

- **Data Acquisition Layer (Cyber Threat Dashboard):** The first layer of the system is responsible for collecting data from multiple sources. This includes static datasets used for training machine learning models as well as real-time threat intelligence feeds obtained from external platforms such as AlienVault OTX and other OSINT sources. The collected data primarily consists of indicators of compromise (IOCs), such as IP addresses and domain names. This layer ensures continuous inflow of both historical and live data required for analysis.
- **Data Preprocessing and feature extraction layer:** Once the raw data is collected, it is passed to the preprocessing layer where cleaning and normalization are performed. This includes handling missing values, standardizing formats, and converting indicators into a consistent structure. The feature extraction module then transforms these indicators into numerical feature vectors using techniques such as domain parsing, statistical analysis, and entropy calculation. These features serve as the foundation for machine learning-based classification.
- **Machine Learning and analysis Layer (CTI):** In this layer, the processed feature vectors are fed into trained machine learning models. Multiple algorithms, including Random Forest, Decision Tree, and Logistic Regression, are used to classify indicators as safe or malicious. The best-performing model, selected through comparative evaluation, is deployed for real-time predictions. This layer forms the core intelligence of the system.
- **Visualization and Dashboard Layer (CTI):** The final layer consists of a web-based dashboard that provides an interactive interface for users. It displays real-time threat data, prediction results, and model performance metrics through visualizations such as charts and graphs. Users can input indicators and receive instant classification results, making the system both user-friendly and practical for real-world

applications .Overall, the architecture ensures a seamless flow of data from acquisition to visualization, enabling efficient, real-time, and intelligent cyber threat detection.

- **Backend and API Layer:** The backend layer is implemented using a Flask-based API that acts as an interface between the machine learning model and external applications. It receives user inputs (IP/domain), processes them through the feature extraction module, and returns prediction results. This layer also handles communication with real-time threat intelligence APIs and manages data flow within the system.

### 6. Expected Outcomes

The proposed Cyber Threat Intelligence System is expected to deliver an intelligent, reliable, and real-time cyber threat detection system by integrating machine learning, feature-based analysis, and live threat intelligence within a unified platform. The system focuses on improving detection accuracy, enabling faster decision-making, and providing meaningful insights through visualization and automation.

- **Improved Decision-Making through Visualization:** The interactive dashboard provides clear visual insights into threat trends, model performance, and real-time data. This helps users and analysts better understand patterns and make informed decisions efficiently.
- **Efficient Model Evaluation and Selection:** By comparing multiple machine learning models using performance metrics such as accuracy, precision, and recall, the system identifies the most effective algorithm for threat detection, ensuring optimal performance.

**Table 1 Expected Outcomes and Description**

Expected Outcome	Description
Threat Detection Accuracy	Machine learning-based classification improves identification of malicious indicators.
Real-Time Analysis	Instant prediction using API integration enables quick threat response.
Unknown Threat Detection	Feature-based approach detects zero-day and obfuscated threats.
Visual Insights	Dashboard provides graphical representation of threats and model performance.

- **Enhanced Threat Detection Accuracy:** The use of advanced feature extraction techniques combined with machine learning algorithms enables accurate classification of indicators such as IP addresses and domain names. This reduces false positives and false negatives, ensuring more reliable identification of malicious activities.
- **Real-Time Threat Analysis:** Integration with a Flask-based API and external threat intelligence sources allows the system to provide instant predictions for incoming indicators. This ensures timely detection of threats and supports quick response in dynamic cybersecurity environments.
- **Detection of Unknown and Evolving Threats:** Unlike traditional systems, the proposed solution leverages statistical and structural features (such as entropy and domain patterns) to detect previously unseen or obfuscated threats. This enhances the system's capability to handle zero-day and evolving cyber attacks.

### Future Scope

Future extensions of the Cyber Threat Intelligence System can significantly enhance its effectiveness by incorporating predictive intelligence, advanced analytics, and scalable cybersecurity frameworks. With the rapid evolution of cyber threats, the system can be extended to move beyond reactive detection toward proactive and autonomous threat management. One promising direction is the integration of predictive threat intelligence, where historical attack patterns and real-time indicator streams are analyzed to forecast potential threats before they occur. By leveraging advanced machine learning and time-series analysis, the system can

identify suspicious trends and generate early warnings, improving response time and reducing the risk of cyber incidents. Another important area of future work is cross-platform and device-level expansion. While the current system operates as a web-based dashboard, future implementations can extend support to mobile applications, enterprise security systems, and IoT environments. Integration with endpoint devices and network monitoring tools would allow richer data collection and more accurate threat detection across diverse environments. The Cyber Threat platform presents substantial opportunities for future enhancement:

- **Predictive Threat Monitoring:** Leveraging historical and real-time data to anticipate cyber threats and generate early alerts before attacks occur.
- **Cross-Platform and Device Support:** Extending support to mobile devices, enterprise systems, and IoT environments for improved data collection and usability.
- **Scalable Group Coordination:** Supporting large-scale deployments through cloud-based architecture and distributed processing techniques.

**Table 2 Future Work and Description**

Future Work	Description
<b>Predictive Threat Monitoring</b>	Anticipating cyber attacks using historical patterns and real-time data.
<b>Cross-Platform Expansion</b>	Supporting mobile apps, enterprise tools, and IoT-based systems.
<b>Smart Infrastructure Integration</b>	Leveraging network analytics and security frameworks for adaptive response.
<b>Large-Scale Scalability</b>	Enabling deployment across enterprise and cloud environments efficiently.

### Conclusion

The Cyber Threat Intelligence System presents a comprehensive and intelligent approach to modern

cyber threat detection by integrating machine learning, feature-based analysis, and real-time threat intelligence into a unified platform. The project successfully addresses key limitations of traditional cybersecurity systems, such as reliance on signature-based detection, lack of real-time responsiveness, and absence of integrated analysis frameworks. By leveraging advanced feature extraction techniques, the system is capable of transforming raw indicators of compromise—such as IP addresses and domain names—into meaningful representations for accurate classification. The use of multiple machine learning models and their comparative evaluation further strengthens the reliability and effectiveness of the solution. Additionally, the integration of a backend API enables real-time prediction, while the interactive dashboard enhances usability through clear visualization of threat data and model performance. The proposed system not only improves detection accuracy but also enhances the ability to identify previously unseen and evolving threats. Its modular and scalable architecture ensures adaptability for real-world deployment, making it suitable for various cybersecurity applications.

In conclusion, this project demonstrates the potential of combining data-driven intelligence with real-time systems to create a robust and efficient cyber threat detection platform. It lays a strong foundation for future advancements in predictive analytics, scalable deployment, and intelligent cybersecurity solutions.

### Reference

- [1]. Santos, P., et al. (2025). A Systematic Review of Cyber Threat Intelligence Approaches. PMC (PubMed Central).
- [2]. Kumar, A., et al. (2025). A GIS-Based Multi-Criteria Decision Analysis Framework Using AHP for Land Suitability Assessment. Elsevier GIS Journal.
- [3]. Devi, S., & S. (2025). Cloud Security Automation Through Symmetry: Threat Detection & Response. Security Automation Research.
- [4]. Khan, M., et al. (2025). Cyber Threat Intelligence: A Systematic Review. IEEE Access.
- [5]. Patel, R., & Roy, S. (2025). Machine Learning in Threat Detection Systems. Patel & Roy

- Publications.
- [6]. Johnson, T., et al. (2025). Integrating OSINT for Real-Time Cyber Threat Analysis. Elsevier.
- [7]. Ahmed, M., & Zhou, L. (2025). Blockchain Enhanced Threat Attribution. ACM Digital Library.
- [8]. PMC Research Authors. (2024). Cyber Threat Intelligence Adoption in Healthcare Security Operations. PMC Article.
- [9]. Li, H., & Fernandez, J. (2024). Predictive Analytics for Cyber Risk Assessment. ScienceDirect.
- [10]. Gupta, A., & Mehra, S. (2024). Cross-Domain Threat Correlation Using Graph Neural Networks. MDPI.
- [11]. Sharma, R., & Lee, J. (2024). Threat Profiling Using Deep Learning. SpringerLink.
- [12]. Wang, X., et al. (2024). Federated Learning for Privacy-Preserving CTI. Elsevier.
- [13]. A. Wu, Y., et al. (2024). A Hybrid Knowledge Graph–LLM Framework for Cyber Threat Intelligence Credibility Assessment. AI and Cybersecurity Journal.
- [14]. Park, T., et al. (2024). Cross-Domain Threat Correlation Using Graph Neural Networks. IEEE Xplore.
- [15]. Zhou, L., et al. (2024). Advancing Cybersecurity: A Comprehensive Review of AI-Driven Detection Techniques. ScienceDirect.
- [16]. Costa, R., et al. (2024). Cyber Threat Data Normalization Framework. Taylor & Francis.
- [17]. Fernandez, P., et al. (2024). OTX-Integrated Threat Sharing and Automation. ResearchGate.
- [18]. P. Kumar, S., & Das, P. (2024). Hybrid SIEM–SOAR Systems for Threat Management. Springer.
- [19]. Verma, A., et al. (2024). Deep Neural Architectures for Phishing Detection. MDPI.
- [20]. Morgan, L., & Patel, R. (2024). Threat Intelligence Sharing Using STIX/TAXII Automation. IEEE Access.
- [21]. Chen, Y., et al. (2024). Machine Learning for Threat Indicator Prioritization. Computers & Security (Elsevier)
- [22]. H. Rodrigues, M., et al. (2023). Automated CTI Enrichment Using OSINT and Dark Web Mining. ACM Digital Threats.
- [23]. H Alvarez, J., & Smith, K. (2023). Deep Learning for Malware Behavior Profiling. Journal of Cybersecurity (Springer).
- [24]. H Park, T., & Liu, H. (2023). GNN-Based Attack Path Identification for Cyber Threat Intelligence Systems. IEEE Transactions on Information Forensics and Security.
- [25]. H Abedin, M., et al. (2023). A Multi-Source Fusion Framework for Heterogeneous Cyber Threat Intelligence. MDPI.
- [26]. G Mahmoud, S., et al. (2023). Zero-Day Vulnerability Prediction Using Cyber Threat Intelligence Signals. Wiley – Security & Communication Networks.
- [27]. Reddy, S., & Chan, P. (2023). IoT Threat Intelligence Aggregation Framework. ACM.
- [28]. H Omia, E., et al. (2023). Real-Time Malware Classification Using Ensemble Machine Learning. Verma et al. Publication.
- [29]. Kim, H., & Alvarez, L. (2023). SOC Automation Through LLMs. ScienceDirect.
- [30]. Oliveira, R., et al. (2023). Comparative Analysis of Threat Intelligence Standards (STIX, TAXII, MISP). Elsevier.
- [31]. Kumar, V., & Jensen, R. (2023). Improving SOC Efficiency via CTI-Driven Automation. IEEE Security & Privacy.
- [32]. Abadi, M., et al. (2023). Vulnerability Exploitation Prediction Using Machine Learning. Wiley.
- [33]. Ahmed Lee, J., & Park, S. (2023). Digital Twin-Based Cyber Threat Simulation. SpringerLink.
- [34]. Singh, P., et al. (2023). CTI Visualization Using Dashboard Analytics. IEEE Access.
- [35]. Ferreira, L., & Zhou, D. (2023). Blockchain-Based Secure Threat Intelligence Exchange. Elsevier – Future Generation Computer Systems.
- [36]. Yang, S., et al. (2023). Quantum-Inspired Threat Prediction Model. IEEE Xplore.
- [37]. Zhao, Q., & Lin, F. (2023). Cyber Threat Hunting via Reinforcement Learning. Springer.