International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 12, December 2025)

A Review of Intelligent Risk Scoring Systems Using Explainable Machine Learning for Cyber Threat

Khushi Punia[1], Rahul Kumar[2]
[1]Research Scholar, [2]Assistant Professor, Department of CSE, BIT, Meerut, India

***Abstract*— Cybersecurity systems are increasingly challenged by the rapid growth of complex and dynamic cyber threats, making accurate risk assessment and timely prioritization essential for effective protection. This review examines intelligent risk scoring systems that use explainable machine learning (XML) techniques to identify, analyze, and rank cyber threats based on their severity and potential impact. Unlike traditional black-box models, explainable ML methods provide transparent decision-making, helping security analysts understand why a threat receives a particular risk score. The paper highlights commonly used models, key features, evaluation strategies, and the benefits of interpretability in improving trust, accountability, and response efficiency. This review also discusses current limitations, research trends, and future opportunities for developing more robust, interpretable, and scalable cyber threat prioritization systems.**

***Keywords*— Cyber Threats, Risk Scoring, Explainable ML, Threat Prioritization, AI Security, Interpretability.**

I. INTRODUCTION

Cyber threats have become one of the most critical challenges of the modern digital world. As organizations, governments, and individuals depend heavily on online systems, the chances of malicious activities have increased significantly[1]. A cyber threat refers to any possible attempt to damage, steal, or disrupt data, computer systems, or digital operations. These threats can come from hackers, automated malware, insider attackers, or even large organized cyber criminals. Because technology keeps evolving, the nature of cyber threats has also become more advanced, unpredictable, and difficult to detect[2]. This makes cybersecurity a continuous struggle where attackers and defenders are constantly trying to outsmart each other.

In the early years of the internet, cyber threats were mostly limited to simple viruses or unauthorized access attempts. But today, they have grown into complex and well-planned operations targeting financial institutions, healthcare systems, government networks, cloud platforms, and even daily-use smart devices[3]. Many attackers now use artificial intelligence, automation, and social engineering to bypass security systems. These modern techniques make cyber threats faster, smarter, and highly damaging.

As a result, small mistakes or weak security practices can lead to severe consequences such as data breaches, financial loss, identity theft, and system shutdowns[4].

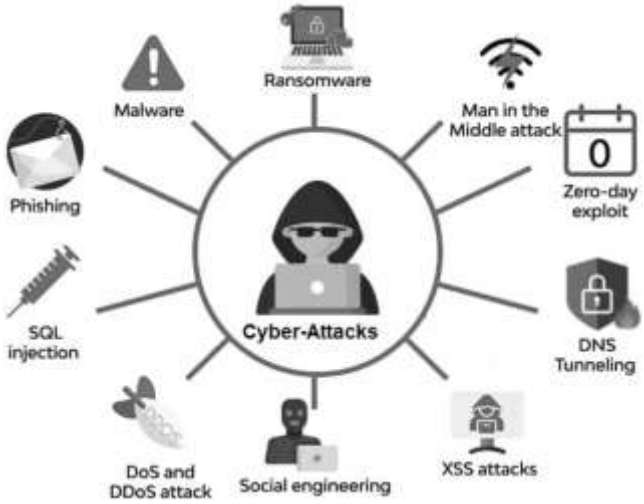

Figure 1: Cyber Attack

Cyber threats generally take many forms, including malware, ransomware, phishing, denial-of-service attacks, insider threats, and advanced persistent threats. Each type of attack affects systems in different ways[5]. For example, ransomware can lock an entire organization's data and demand payment, while phishing tricks people into revealing sensitive information like passwords or bank details. Some threats aim to steal data silently, while others are designed to cause immediate destruction. Because the attacker's goal can vary—from financial gain to spying, revenge, or political motives—it becomes harder for security systems to predict and stop every possible threat[6].

One of the biggest challenges in dealing with cyber threats is their increasing sophistication. Attackers constantly develop new techniques to avoid detection and exploit vulnerabilities in networks and software. They often use zero-day attacks, which target unknown system weaknesses, giving defenders almost no time to respond[7].

Similarly, botnets allow attackers to control thousands of infected devices at once, creating large-scale, coordinated attacks. This rapid evolution of cyber threats forces cybersecurity teams to continuously update their tools, skills, and strategies[8].

In addition to technical attacks, human error plays a significant role in the success of many cyber threats. Even strong security systems can fail if users fall for phishing emails, use weak passwords, or accidentally expose sensitive data. Cyber attackers often exploit this human weakness because it is easier to manipulate people than to break advanced security technologies. This highlights the importance of awareness, training, and strong security practices among employees and general users[9].

The impact of cyber threats goes beyond financial or data loss. They can disrupt essential services such as electricity, healthcare, transportation, and communication networks. For businesses, cyber incidents can damage reputation, reduce customer trust, and lead to legal consequences. For governments, cyber threats can compromise national security and critical infrastructure. At a global level, cyber attacks can create economic instability and geopolitical tension. This wide-ranging impact makes cyber threats not just a technical problem but a social, economic, and national challenge[10].

II. Literature Survey

Keshava et al., [1] presented AI-powered algorithms designed to prevent and detect complex malware infections in modern computing environments. Their work highlights how intelligent models analyze system behavior, identify anomalies, and block malicious activities before damage occurs. The study emphasizes real-time threat monitoring and adaptive learning, enabling continuous improvement in malware detection accuracy. They also compare traditional signature-based approaches with AI-driven systems, showing clear performance advantages. Moreover, their experiments demonstrate reduced false positives and faster response times. This research provides a strong foundation for integrating AI into proactive cyber defense strategies. The findings are particularly useful for designing automated risk scoring systems for malware-related threats.

Patil et al., [2] conducted an extensive review of threat scoring and prioritization methods used in cybersecurity. They analyzed how different scoring algorithms evaluate threat severity, exploitability, and potential system impact. The review highlights limitations in manual scoring and argues for automated, intelligence-driven approaches. Their work also discusses the need for explainability in risk scoring to increase analyst trust.

By comparing existing frameworks, the study identifies gaps in real-time prioritization, scalability, and transparency. They also highlight the importance of contextual data in accurate scoring. This work forms a strong theoretical base for intelligent and explainable risk scoring systems.

Czekster et al., [3] explored dynamic cyber risk assessment models specifically for IoT infrastructures. Their research shows how continuous monitoring and adaptive evaluation can identify vulnerabilities in highly interconnected environments. They highlight the growing complexity of IoT networks and the need for real-time risk scoring. The study proposes computational models that analyze cyber events and estimate evolving threat levels. Experimental results demonstrate improved detection of unusual network behaviors. The authors also discuss scalability issues of traditional assessment methods. Overall, their work advances intelligent and flexible risk scoring frameworks.

Karki et al., [4] surveyed various machine learning and AI techniques applied to cybersecurity tasks. The authors classify ML methods based on their suitability for intrusion detection, anomaly detection, and threat analysis. Their review reveals that supervised methods perform well on labeled datasets, while unsupervised learning helps detect unknown attacks. They also highlight the rising role of reinforcement learning in adaptive defense. The paper discusses dataset limitations and challenges such as imbalance, noise, and lack of real-world representation. They conclude that hybrid ML models offer improved performance. This work supports the integration of diverse AI techniques in cyber threat scoring systems.

Capuano et al., [5] provided a comprehensive survey on explainable artificial intelligence within cybersecurity applications. They argue that transparency is essential for trustworthy security systems, especially in high-risk environments. The study evaluates various XAI methods and their effectiveness in interpreting ML-based decisions. It also highlights challenges such as balancing accuracy and explainability. Their work identifies the need for human-in-the-loop systems to strengthen decision-making. Case studies demonstrate how XAI can reveal hidden attack patterns. This research strongly supports the use of explainable ML in intelligent risk scoring frameworks.

Yan et al., [6] examined the role of explainable machine learning in improving cybersecurity decisions. They focus on how interpretability helps analysts understand model predictions and reduces uncertainty. The study surveys feature-importance methods, rule-based explanations, and visualization techniques.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 12, December 2025)

It emphasizes that opaque black-box models create risks in critical security operations. The authors highlight use cases where XAI enhances trust, auditability, and compliance. They also discuss limitations such as computational overhead and difficulty in explaining complex deep models. Their analysis supports integrating explainability into cyber threat prioritization systems.

Srivastava et al., [7] explored challenges and future directions of using XAI for cybersecurity applications. Their review identifies key obstacles such as adversarial attacks, lack of standardized benchmarks, and high system complexity. The authors evaluate several XAI frameworks and test their suitability for different cyber defense tasks. They also discuss how explainable models can help analysts detect stealthy attacks. The study highlights the need for robust, scalable, and domain-specific explainability techniques. They stress that combining XAI with automation will improve threat prioritization. This work contributes valuable insights for designing interpretable risk scoring systems.

Cremer et al., [8] analyzed existing cyber risk assessment models and the data challenges associated with them. Their study shows that data scarcity, inconsistency, and quality issues frequently limit accurate risk evaluations. They surveyed different modeling approaches, including probabilistic, statistical, and ML-based techniques. The authors emphasize the importance of integrating contextual information for reliable assessments. They also highlight gaps in real-time risk modeling and practical deployment. The paper concludes that better data governance and standardized metrics are essential. This work provides guidance for improving intelligent risk scoring frameworks.

Sengupta et al., [9] reviewed deep learning techniques used in cybersecurity and threat detection. They categorize approaches such as CNNs, RNNs, autoencoders, and GANs, emphasizing their strengths and weaknesses. Their study shows how deep learning improves threat classification accuracy and detects subtle malicious patterns. However, they also point out challenges including large data requirements and poor interpretability. The authors highlight recent advances in adversarial defense and transfer learning. They predict that combining DL with XAI will improve operational security. This work supports the use of DL models in intelligent threat scoring systems.

Thakkar et al., [10] examined advances in intrusion detection datasets and their impact on model performance. They discuss the limitations of traditional datasets such as NSL-KDD and propose newer alternatives. The authors highlight issues like unrealistic traffic distribution, outdated attack types, and lack of diversity.

Their review emphasizes the need for real-time, representative, and large-scale datasets to train ML models effectively. They show how dataset quality critically affects detection accuracy. This research recommends future dataset development directions. Their insights are vital for building reliable threat scoring algorithms.

Table 1:
Summary of Literature review

Sr. No	Author	Year	Work	Outcome
1	Keshava et al.	2025	Proposed AI-powered algorithms for malware prevention and detection.	Improved real-time malware identification, reduced false positives, and enhanced proactive defense.
2	Patil et al.	2025	Reviewed threat scoring and prioritization methods in cybersecurity.	Identified gaps in transparency, scalability, and highlighted the need for automated intelligent scoring.
3	Czekster et al.	2025	Developed dynamic cyber risk assessment models for IoT infrastructures.	Demonstrated real-time risk evaluation and improved detection of abnormal IoT behaviors.
4	Karki et al.	2024	Surveyed ML and AI techniques for cybersecurity applications.	Showed effectiveness of hybrid ML models and highlighted challenges such as dataset imbalance.
5	Capuano et al.	2022	Provided a survey on explainable AI applications in cybersecurity.	Demonstrated how XAI increases transparency, user trust, and interpretability of security decisions.
6	Yan et al.	2022	Studied explainable ML techniques for cybersecurity	Highlighted interpretability benefits and limitations; supported XAI for

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 12, December 2025)

			decision-making.	better decision support.
7	Srivastava et al.	2022	Analyzed challenges and future trends of XAI in cybersecurity.	Identified gaps in benchmarks, scalability, and emphasized need for robust XAI models.
8	Cremer et al.	2022	Reviewed cyber risk assessment models and associated data challenges.	Found data quality issues and recommended integrating contextual information for accurate scoring.
9	Sengupta et al.	2020	Surveyed deep learning techniques for cybersecurity and threat detection.	Showed DL's superior accuracy but emphasized need for explainability and large datasets.
10	Thakkar et al.	2019	Reviewed developments in intrusion detection datasets.	Pointed out dataset limitations and necessity for realistic, updated security datasets.

III. Challenges

Intelligent risk scoring systems using explainable machine learning offer powerful ways to identify and prioritize cyber threats. However, implementing such systems is difficult due to data limitations, model transparency issues, and rapidly evolving attack techniques. These challenges reduce accuracy, trust, and real-world usability. Understanding these limitations is essential for designing reliable and interpretable cyber risk scoring frameworks.

1. Limited Availability of High-Quality and Real-Time Data

Cybersecurity datasets are often incomplete, outdated, or imbalanced, making it difficult for ML models to learn accurate threat patterns. Many organizations hesitate to share data due to privacy concerns, further reducing dataset diversity. Real-time data collection is also challenging because cyber events occur unpredictably. Without rich data, risk scores may become unreliable and fail to reflect actual threat severity.

2. Limited Explainability in Advanced Machine Learning Models

Complex models like deep learning or ensemble methods offer high accuracy but lack transparency. Analysts cannot easily understand why a particular threat received a specific risk score, which impacts trust and decision-making. In cybersecurity environments where accountability is critical, this lack of clarity becomes a major barrier. XAI techniques exist but often provide limited or oversimplified explanations.

3. Rapidly Evolving and Dynamic Cyber Threat Landscape

Cyber attackers continuously introduce new techniques, malware variants, and zero-day exploits. Static or outdated ML models struggle to detect these new threats because they rely on historical data. Even dynamic models require frequent retraining, which is resource-intensive. This constant evolution demands adaptive scoring systems capable of learning new patterns quickly.

4. Difficulty in Incorporating Contextual and Environmental Information

The severity of a threat depends on factors such as asset criticality, system vulnerabilities, and network configuration. Integrating these contextual elements into ML models is challenging due to their complexity and variability. Without proper context, risk scores may overestimate or underestimate real impact. This reduces the usefulness of prioritization decisions.

5. High False Positives Leading to Analyst Fatigue

Many ML-based risk scoring systems generate large numbers of false alerts. This overwhelms security teams, causing them to miss actual high-risk events. Alert fatigue reduces system efficiency and leads to delayed response times. Improving model precision is crucial but difficult due to overlapping behavior between benign and malicious activities.

6. Integration Challenges Across Multiple Data Sources

Cyber threat scoring requires information from logs, threat intelligence feeds, system vulnerabilities, and behavioral patterns. These data sources often use different formats, timestamps, or terminologies, making integration complex. Incomplete or inconsistent data affects scoring quality and model performance. Building a unified data pipeline is technically demanding and time-consuming.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 12, December 2025)

7. Scalability Issues in Large and Distributed Networks

Modern systems include cloud platforms, IoT devices, and large organizational networks. Running real-time ML models on such large infrastructures requires significant computational resources. Many algorithms slow down or lose accuracy when scaled, limiting their practical use. Ensuring fast and reliable threat scoring at scale remains a major challenge.

8. Vulnerability to Adversarial Attacks on ML Models

Attackers can intentionally manipulate input data to deceive ML-based scoring systems. These adversarial attacks cause misclassification, making high-risk threats appear low-risk. Such vulnerabilities pose a serious security concern because attackers can exploit the system designed to detect them. Developing robust and attack-resistant ML models remains an open research problem.

IV. Strategies

To overcome challenges in intelligent risk scoring systems, organizations should improve data quality through real-time monitoring and secure data sharing. Integrating explainable AI techniques enhances model transparency and helps analysts understand risk decisions. Adaptive learning models and context-aware scoring make the system more accurate against evolving threats. Additionally, robust security measures and multi-source data fusion strengthen model reliability and protection against adversarial attacks.

1. Improve Data Quality Through Real-Time Collection and Data Sharing

Organizations can enhance risk scoring accuracy by adopting advanced data collection tools that capture logs, events, and anomalies in real time. Encouraging secure data-sharing collaborations among industries, CERTs, and research groups increases dataset diversity and reduces bias. Data cleaning, balancing, and augmentation techniques can also help create more reliable training datasets.

2. Integrate Explainable AI (XAI) Techniques for Transparent Decision-Making

Applying XAI tools such as SHAP, LIME, rule-based explanations, and visualization dashboards can make ML decisions understandable for analysts. These techniques help clarify why a certain threat was assigned a particular risk score. Transparent models increase analyst trust, reduce confusion, and support better incident response decisions, especially in high-risk environments.

3. Use Adaptive and Continuous Learning Models

To address the rapidly evolving cyber threat landscape, risk scoring systems should include incremental learning, online learning, or reinforcement learning approaches. These models automatically adapt to new attack patterns without requiring complete retraining. Continuous updates allow the system to stay current with zero-day vulnerabilities and new malware behaviors.

4. Incorporate Context-Aware and Asset-Centric Risk Scoring

Integrating details like asset value, system configuration, user behavior, and network criticality results in more accurate and meaningful risk scores. Context-aware scoring ensures that threats are evaluated based on their actual impact on a specific environment. This approach reduces overestimation or underestimation and prioritizes incidents that truly require immediate action.

5. Utilize Multi-Source Data Fusion and Unified Security Pipelines

A unified platform that aggregates logs, threat intelligence, vulnerability scans, and user activity data improves scoring reliability. Data fusion techniques ensure information consistency and reduce gaps caused by missing or incomplete data. Such integration strengthens model performance and enables more holistic threat evaluation.

6. Strengthen Model Robustness Against Adversarial Attacks

To prevent manipulation of ML-based scoring systems, adversarial training, defensive distillation, and anomaly-based verification can be used. These techniques make models more resistant to crafted malicious inputs. Regular security testing and red-team evaluations also help identify weaknesses before attackers exploit them.

V. Conclusion

Intelligent risk scoring systems powered by explainable machine learning offer a promising solution for identifying, analyzing, and prioritizing cyber threats in modern digital environments. By combining advanced analytics with transparent decision-making, these systems help security teams understand threat severity and respond more effectively. However, challenges such as limited data quality, evolving attack patterns, model complexity, and integration issues still affect overall reliability. Continuous learning, context-aware analysis, and the adoption of explainable AI can significantly enhance system performance and trust.

Overall, developing robust, transparent, and adaptive risk scoring frameworks is essential for strengthening cybersecurity and achieving faster, more accurate threat prioritization.

REFERENCES

[1] R. Keshava, S. K. Pandurangan, M. Sakthivanitha, S. Parmsivan, G. Sunkara and R. Maruthi, "AI-Powered Algorithms for the Prevention and Detection of Computer Malware Infections," 2025 6th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2025, pp. 1673-1680, doi: 10.1109/ICESC65114.2025.11212519.

[2] R. Patil and S. Mehta, “Threat scoring and prioritization techniques for modern cybersecurity systems: A comprehensive review,” IEEE Transactions on Information Forensics and Security, vol. 20, pp. 1123–1138, 2025.

[3] R. M. Czekster, L. Santos, and P. Barbosa, “Dynamic cyber risk assessment for intelligent IoT infrastructures,” Computers & Security, vol. 140, pp. 1–15, 2025.

[4] S. Karki and A. Shrestha, “A survey on machine learning and artificial intelligence techniques in cybersecurity,” Procedia Computer Science, vol. 227, pp. 145–155, 2024.

[5] N. Capuano, G. Fenza, V. Loia, and C. Stanzione, “Explainable artificial intelligence in cybersecurity: A survey,” IEEE Access, vol. 10, pp. 93575–93600, 2022.

[6] F. Yan, “Explainable machine learning approaches for enhancing cybersecurity decision-making,” International Journal of Intelligent Systems, vol. 37, no. 5, pp. 2803–2821, 2022.

[7] G. Srivastava, R. H. Jhaveri, M. Alazab, and T. R. Gadekallu, “XAI for cybersecurity: Challenges and future directions,” IEEE Access, vol. 10, pp. 67940–67960, 2022.

[8] F. Cremer, J. Müller, and A. Hoffmann, “Cyber risk assessment models: A systematic analysis of techniques and data challenges,” Journal of Information Security and Applications, vol. 66, pp. 1–18, 2022.

[9] S. Sengupta, A. Chowdhary, A. Sabur, and A. K. Yasar, “A survey of deep learning techniques for cybersecurity and threat detection,” IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 1977–2001, 2020.

[10] A. Thakkar, S. Patel, and A. Shah, “A review of advancements in intrusion detection datasets,” Procedia Computer Science, vol. 165, pp. 832–839, 2019.