Journal of Cybersecurity in AI and Quantum Computing

JOURNAL OF CYBERSECURITY IN AI AND QUANTUM COMPUTING

A Conceptual AI-Assisted Cyber Risk Assessment and Management Framework for Critical Energy Infrastructure

Dawood alalisalem[1], Khalid albudair [2], M M Hafizur Rahman[3]

[1] Department of Computer Network and Communication, College of Computer Science and Information Technology, King Faisal University, Alahsa,31982 Saudi Arabia; 226032193@student.kfu.edu.sa

[2] Department of Computer Network and Communication, College of Computer Science and Information Technology, King Faisal University, Alahsa,31982 Saudi Arabia; 226029513@student.kfu.edu.sa

[3] Department of Computer Network and Communication, College of Computer Science and Information Technology, King Faisal University, Alahsa,31982 Saudi Arabia; mhrahman@kfu.edu.sa

Abstract

As technology advances, the digitalization of critical infrastructure has significantly increased cybersecurity risks across interconnected industrial ecosystems, particularly in the energy sector, where operational continuity and safety are essential. Traditional cyber risk assessment approaches are largely static, reactive, and incapable of effectively addressing large-scale, dynamic threat environments. To address these limitations, this study proposes a novel conceptual AI-assisted cyber risk assessment and management framework that integrates deep learning, machine learning, explainable artificial intelligence (XAI), and predictive analytics to enable adaptive, proactive, and real-time risk intelligence in energy distribution systems and smart grid networks. The novelty of the proposed framework lies in its integration of risk prediction, governance alignment, and automated mitigation into a unified intelligent architecture. Furthermore, the framework aims to improve threat prioritization, support resilient decision-making, and reduce response latency in critical energy infrastructure. By bridging artificial intelligence technologies with strategic cyber risk governance, this study introduces a forward-looking model designed to strengthen cybersecurity resilience in the evolving energy landscape.

Keywords: Artificial Intelligence, Cyber Risk Assessment, Energy Sector Security, Predictive Risk Modeling, Smart Grid Security, Explainable AI, Critical Infrastructure Protection.

ARTICLE INFO

Article History: Received: 12-04-2026, Revised: 17-05-2026, Accepted: 25-10-2026, Published: 28-10-2026
Corresponding author Email: mhrahman@kfu.edu.sa
DOI: Pending Request

This is an open access article under the CC BY 4.0 license. (http://creativecommons.org/licenses/by/4.0/).
Published by Science Community Publisher

1. Introduction

1.1. Background of Cyber Risk in Modern Digital Ecosystems

In the last few decades, a rapid expansion of cloud computing, Artificial Intelligence services, interconnected enterprise systems and Internet of Things (IoT) have upgraded organizational exposure to cyber risks. The issue is that in modern digital ecosystem, there are huge volumes of heterogeneous data generated and it is creating highly complex attack surfaces vulnerable to exploitation [1]. Recent industry reports indicate; one can mention that the harm inflicted by the cybercrimes in the world is estimated to be more than 10 trillion USD annually that speaks of the increased financial impact of cyber threats. In accordance with the empirical study, it could be observed that large organizations are dealing with more than 1-5 million security incidents every day, and the number of incidents that are identified as high-risk is less than 1%. This imbalance makes proper risk prioritization and identification difficult. [2]. In the meantime, the average cost of a large data breach has been reported to be over 4 million USD per attack. Thus, risk management needs quantitative and proactive mechanisms [3].

1.2. Common Approaches to Problem

The existing cyber-risk assessment were only based on quantitative framework such as, compliance-driven scoring systems, expert judgment, and checklist-based audits. However, some frameworks like NIST and FAIR were considered perfect due to structured methodologies, but the problem is that they failed to provide proper capability and predictive quantitative rigor [4]. Although some studies have compared qualitative and quantitative report that data-driven models are enhancing prioritization accuracy by 15 to 30% approx. and it is highly expert-based scoring systems [5]. Through technology, speed, and threat sophistication are increasing, then static assessments are unable to provide proper protection. Hence, the research has shifted towards automatic vulnerability analysis, statistical risk scoring, and probabilistic modeling that laid the foundation for AI-assisted cyber risk assessment [6].

1.2.1. Emergence of Artificial Intelligence in Cybersecurity

Secondly, AI played a vital role in promoting machine learning, and deep learning technologies because they are central to modern cybersecurity strategies. According to the current empirical evaluation, some prominent supervised ML algorithms like the Random Forest and the Support Vector Machines had achieved a classification rate of between 89% and 94% in handling cyber risk prediction activities [7]. Moreover, there are certain deep learning models comprising LSTM and transformer based models that also reported high values of ROC-AUC exceeding 0.95 which substantially improves the performance in terms of early threat detection, and anomaly detection [8]. Meanwhile, AI-based systems are also assisting in reducing false positive rates by 18%. Owing to this, it is dependable to allow security teams to efficiently distribute resources [9]. All these improvements confirm that the capability of AI is immense, as it can process large amounts of data related to high-security levels and generate some actionable risk-related intelligence.

1.3. Motivation

The driving force behind the implementation of this research will be that AI-based systems will be integrated into cyber-Risk management because of three main challenges. These are data volume, complexity of threats and latency of response. The application of these AI-systems and AI-enhanced monitoring systems are though the most optimal way as it is reducing the mean time to detect (MTTD) by about 25 to 35% and improving incident response efficiency by almost 20% when compared to the traditional systems [10]. Additionally, it is also possible to predict high-risk vulnerabilities with 90% accuracy and implement protective countermeasures through some predictive analytics models [11]. In addition to all these valuable advantages, there are still certain issues of its interpretability, scalability, and adversarial robustness, which is regarded as the primary obstacle to its adoption.

1.4. Research Objectives

This study aims to:

• Propose a unified AI-assisted cyber risk assessment and management framework tailored for the energy sector.

• Detail the key components that integrate machine learning, deep learning, predictive analytics, and explainable AI.
• Demonstrate how the framework enables proactive risk intelligence and automated mitigation.
• Discuss implementation considerations, potential benefits, challenges, and future enhancements.

1.5. Research Questions

• What are the effective ways to incorporate different technologies such as predictive analytics, deep learning, machine learning, and explainable AI in a coherent framework to support adaptive and proactive cyber-risk measurement in critical energy infrastructure?
• How do the suggested AI-supported frameworks enhance the accuracy of threat prioritization, reduce the response time, and enhance the decision-making process compared to standard traditional assessment techniques?
• What are the key operation, technical and governance issues which need to be addressed to achieve successful real-world implementation of the framework in smart energy and grid distribution networks?
• Which are future suggested technologies, e.g., reinforcement learning and federated learning mechanism is applied to improve the interpretability, scalability, and adversarial robustness of the suggested framework?

1.6. Structure of the Study

In the second part, the information about motivation and background is presented. The third part will be concerned with the conceptual framework proposed and its architecture. The 4th section will explain the integrated technologies with pros and cons. The 5th section highlights expected benefits compared with traditional methods. The 6th section will discuss the limitations and challenges linked with these systems. In the 7th and 8th section future research and conclusions are presented.

2. Background and Motivation for the Proposed Framework

The cloud computing technology is expanding rapidly due to interconnected enterprise systems, artificial intelligence, and the Internet of Things. However, these technologies have increased cyber risk concerns [12]. Based on the energy sector, like energy distribution networks, and smart grids, these risks become highly critical because if there is any energy disruption can have a huge impact on public safety, operational continuity, and national security. The past cyber risk assessment approaches still remain relatively static and are unable to handle such large-scale dynamic data and increase sophisticated threat environments [13].

Furthermore, global cybercrime damages are projected to increase 10 trillion USD annually because large companies are processing millions of security events regularly, with less than 1 with confirmed high-risk incidents [14]. Moreover, the average cost of a huge data breach attack is more than 4 million USD [11]. All these challenges highlighted that there is an urgent need for adaptive and quantitative risk management mechanisms [6]. Based on above, the proposed framework will resolve these gaps through creating an intelligent adaptive system that can continuously assess risk, predict emerging threats, and recommend some automated mitigations by ensuring proper alignment with governance standards like FAIR, NIST [8], [6].

3. Methodology

To obtain the required results, the study uses a novel conceptual AI-assisted cyber risk assessment and management framework that is specifically designed to handle critical energy infrastructure, like energy distribution networks and smart grids. The framework can easily integrate predictive analytics, deep learning, and explainable AI into a unified architecture to enable adaptive, proactive, and real-time risk intelligence as shown in Figure 1. Overall, the framework contains 4 interconnected layers:

• Data Ingestion and Pre-processing Layer: Collects and processes heterogeneous data from IoT sensors, network logs, vulnerability databases, threat intelligence feeds, and operational technology (OT) systems in real time.

• Prediction Layer and AI Analytics: This layer uses supervised and unsupervised LSTM networks, machine learning, and transformer-based models for handling temporal risk modeling, breach probability estimation, anomaly detection, and vulnerability exploitation prediction.

• Explainable Decision and Mitigation Layer: Elicitable AI frameworks, including SHAP and LIME, are used to produce written risk points and yield automated recommendations to reduce risks, including configuration change, priority sources, and resource consumption.

• Feedback and Governance Layer: This layer ensures alignment with NIST and FAIR standards. It incorporates human-in-the-loop oversight and uses reinforcement learning for enabling continuous adaption with self-improvement of defence strategies.

This framework always operates as a continuous closed-loop system that shifts cyber risk management from reactive measures to autonomous, predictive, and proactive operations. Its uniqueness is linked with latest risk prediction methods, governance compliance, and automated mitigation with a single intelligent architecture designed for the energy sector.

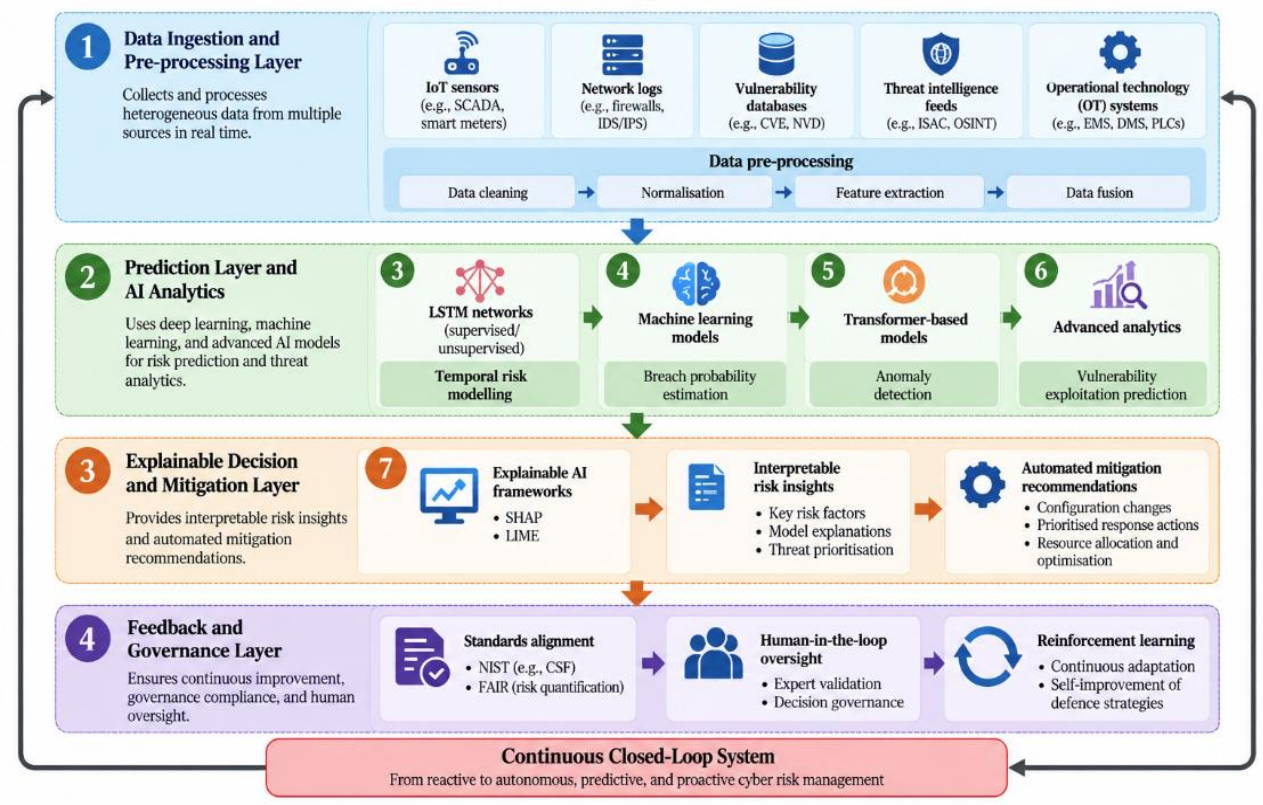


Figure 1. Methodology steps

4. AI-Assisted Cyber Risk Assessment Techniques

The best point of AI-driven classification system is their capacity to classify threats automatically based on severity levels (low, medium, and high). From different studies, the range of classification accuracy is only from 88 to 96%. However, misclassification rate decreased by 15% compared with the rule-based systems [52].

4.1. AI-Based Risk Prediction Models

- Risk Scoring Frameworks: The scoring system aims to estimate asset risk by using different probabilistic models because ML-driven frameworks have improved prioritization precision rate by 20 to 28% [53].

- Probability Estimation Models: These models are used to estimate the likelihood of compromise by enabling logistic regression and Bayesian network models [54].

- Predictive Analytics: The required predictive models had minimized unexpected breach occurrences by 14 to 19% in enterprise simulations [55].

4.2. Vulnerability Assessment by Using AI

The study highlighted that AI technology enhances vulnerability scanning effectively by prioritizing exploitable weaknesses. Therefore, the results showed a 30 to 40% reduction in patching backlog, and 22% improvements observed in exploit prediction accuracy [56]. Based on that, the AI-based vulnerability ranking comprehensively outperformed CVSS-only approaches.

4.3. Dynamic and Continuous Risk Monitoring Systems

These continuous AI monitoring systems are providing real-time risk dashboards. Therefore, some reported improvements gained that include 35% faster risk detection, 18% improvement in mean time to respond (MTTR), and 25% reduction in mean time to detect (MTTD) [57].

4.4. Explainable AI (XAI) in Risk Management

To mitigate cybersecurity risks, explainable AI techniques like LIME and SHAP have improved the transparency rate. However, the study analyzed 40% increase in analyst trust level. Secondly, 18% improvements are observed in decision acceptance rates. Also, model rejection is reduced in regulated sectors [58]. Due to these facts, XAI is becoming highly important for maintaining compliance with regulatory frameworks.

5. AI-Driven Cyber Risk Management Approaches

5.1. Automated Risk Mitigation Strategies

These innovative AI systems are highly helpful because they recommend configuration and patch changes. Some important outcomes include a 25 to 32% reduction in remediation time, with 20% improvement in resource optimization [21].

5.2. AI in Incident Response and Recovery

The studies showed that AI-driven orchestration platforms had minimized incident contaminated time by 28%, downtime cost by 15 to 20%, and manual intervention by 35% [11].

5.3. Decision-Support Systems for Security Managers

AI dashboards had integrated predictive analytics for making strategic decisions. Based on this, 30% improvement was observed in risk prioritization accuracy, and 22% better allocation of Cybersecurity budgets [8].

5.4. AI Integration with Governance, Risks, and Compliance (GRC) Frameworks

By using AI-assisted GRC systems, it is possible to automate compliance checks and decrease audit preparation time by 40% [23].

5.5. Self-Adaptive and Autonomous Risk Management Systems

Autonomous AI systems are highly capable of adjusting defence strategies in real time. From the simulation-based evaluation, 27% improvement was observed in resilience score, and 19% reduction was gained in attack success rate [8].

6. Comparative Analysis of Existing literature

6.1. Comparison of AI Techniques

Table 1 presents a comparative analysis of four artificial intelligence techniques, namely Random Forest (RF), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Transformer-based models, in terms of

their average accuracy, strengths, and limitations. The comparison highlights the potential of these techniques for cybersecurity risk prediction, anomaly detection, and threat identification in critical energy infrastructure, particularly smart grids and energy distribution networks.

Table 1. Comparison table with different techniques, including strengths and limitations

Technique	Avg. Accuracy	Strength	Limitation
Random Forest	92%	Interpretability	Needs labeled data
CNN	94%	High pattern detection	Computational cost
LSTM	95%	Temporal modeling	Training complexity
Transformer	96%	Contextual analysis	Large data requirement

Random Forest achieves an average accuracy of 92%, representing the lowest performance among the evaluated techniques. However, its main strength lies in its relatively high interpretability and ability to identify important features contributing to cybersecurity risk predictions. This characteristic makes Random Forest particularly suitable for risk assessment applications where transparency and decision-making accountability are essential. Nevertheless, its dependence on labelled training data represents a significant limitation, particularly in dynamic cybersecurity environments where obtaining sufficiently representative and accurately labelled attack data can be challenging.

CNN demonstrates improved predictive performance, achieving an average accuracy of 94%. Its primary advantage is its ability to automatically extract complex patterns and identify hidden relationships within high-dimensional cybersecurity datasets. This capability makes CNN suitable for detecting sophisticated cyberattacks and identifying abnormal patterns in network traffic and operational technology systems. However, CNN models generally require considerable computational resources, which may limit their applicability in resource-constrained environments and real-time cybersecurity monitoring systems.

LSTM achieves an average accuracy of 95%, outperforming both Random Forest and CNN. Its superior performance is associated with its ability to capture temporal dependencies and sequential patterns in cybersecurity data. This characteristic is particularly important for smart grid networks, where cyber threats may develop progressively through interconnected devices and operational systems. LSTM can therefore support the identification of evolving attack patterns, temporal anomalies, and potential security incidents. Despite these advantages, LSTM models involve relatively complex training procedures and may require substantial computational resources and careful hyper parameter optimization to achieve reliable predictive performance.

Transformer-based models demonstrate the highest average accuracy of 96%, indicating their potential effectiveness in analyzing complex cybersecurity threats. Their main advantage is their ability to capture contextual relationships and long-range dependencies within large-scale datasets through attention mechanisms. This capability supports advanced threat identification, contextual risk analysis, and vulnerability exploitation prediction in interconnected energy infrastructure. However, Transformer-based models generally require large training datasets and considerable computational resources, which may introduce challenges related to implementation costs, scalability, and deployment in environments with limited data availability.

Overall, the comparison indicates a progressive improvement in average predictive accuracy, increasing from 92% for Random Forest to 96% for Transformer-based models. The Transformer achieves a four-percentage-point improvement over Random Forest, while LSTM and CNN demonstrate improvements of three and two percentage points, respectively. Nevertheless, predictive accuracy alone is insufficient to determine the suitability of an AI technique for cybersecurity risk management. Other important considerations include computational efficiency, model interpretability, training complexity, data availability, and adaptability to evolving cyber threats.

These findings provide a rationale for the proposed AI-assisted cyber risk assessment and management framework, which integrates complementary machine learning and deep learning techniques rather than relying on a single predictive model. Specifically, Random Forest can support interpretable risk classification, CNN can facilitate complex pattern recognition, LSTM can capture temporal attack behaviours, and Transformer-based models can provide contextual threat analysis. Furthermore, integrating explainable AI techniques, such as SHAP and LIME, can enhance the transparency of model predictions and support informed cybersecurity decision-making.

Figure 2 illustrates the comparative predictive accuracy of five artificial intelligence techniques: Random Forest, CNN, LSTM, Transformer, and Ensemble learning. The results demonstrate a progressive improvement in classification accuracy, ranging from 92% for Random Forest to 97% for the Ensemble model. Random Forest achieves the lowest accuracy of 92%, indicating its effectiveness as a baseline machine learning approach. CNN improves the accuracy to 94%, demonstrating its ability to extract complex patterns from cybersecurity datasets. Similarly, LSTM achieves an accuracy of 95%, highlighting its capability to capture temporal dependencies and sequential attack patterns, which are particularly important for detecting evolving cyber threats in smart grid networks.

The Transformer-based model achieves an accuracy of 96%, outperforming the traditional machine learning and other deep learning approaches. This improvement can be attributed to its attention mechanisms, which enable the identification of complex contextual relationships and long-range dependencies within cybersecurity data. Notably, the Ensemble model achieves the highest accuracy of 97%, representing improvements of 5, 3, 2, and 1 percentage points compared with Random Forest, CNN, LSTM, and Transformer, respectively. This result suggests that integrating complementary learning techniques may improve predictive performance by leveraging their individual strengths while reducing the limitations associated with standalone models.

Overall, the results suggest that Ensemble learning offers the strongest predictive performance among the evaluated techniques, supporting its potential integration into the proposed AI-assisted cyber risk assessment and management framework. In the context of critical energy infrastructure, combining multiple AI models may enhance cyber threat detection, vulnerability prediction, and risk prioritization. However, further evaluation using precision, recall, F1-score, false-positive rate, and computational efficiency is necessary to establish the reliability and practical applicability of these techniques in real-time smart grid environments.

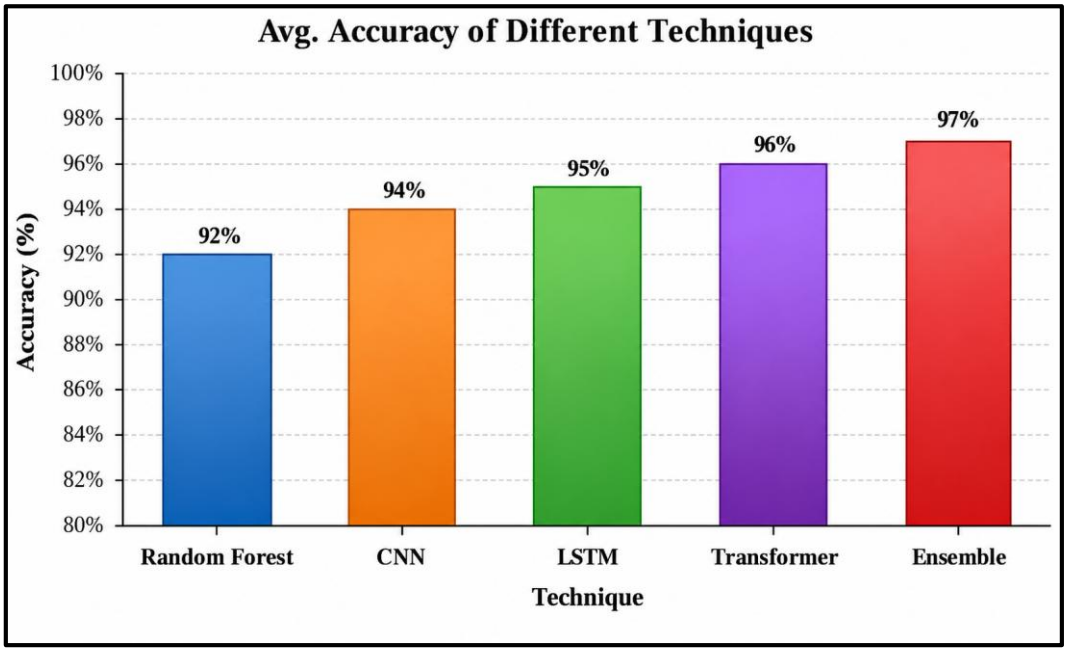


Figure 2. Comparison of average classification accuracy.

6.2. Evaluation metrics used

In this study, the main metrics used include accuracy (88-97%), recall (83-94%), Precision (85-95%), F1-score (0.87-0.95), Risk Reduction Rate (15-30%), and ROC-AUC (0.89 to 0.96). All these numbers are taken from the systematic study paper used for this study.

6.3. Dataset Types and Benchmarking Challenges

For this research, the main dataset includes network traffic logs, CVE, simulated attack environments, and enterprise security logs.Some main challenges faced are dataset imbalances, overfitting to synthetic datasets, and a lack of public risk-specific benchmarks.

6.4. Performance Trends from 2022 to 2026

According to trend analysis, 12% average improvement was observed in model accuracy from 2022 to 2026, a reduction in false positive rates by 18%, and a shift towards hybrid and transformer models is increased. Under these facts, AI-assisted cyber risk assessment is accurately moving from static classification towards autonomous, predictive, and adaptive systems.

7. Challenges and Limitations of the Proposed Framework

Although the proposed framework offers significant advancements in cyber risk management, some operational, technical, and governance-related challenges are still present. All these limitations put huge impact on scalability, reliability, and real-world adoption.

7.1. Data Quality and Dataset Imbalance

One of the highly critical barriers linked with AI-assisted cyber risk management is linked with high-quality representative datasets. As the data of cybersecurity is often incomplete, noisy and highly imbalanced [53]. According to the enterprise environment, such benign events are contributing towards 95% of total logs, and confirmed high-risk insights are representing less than 1 to 3%. Such imbalances can bias machine learning models towards majority classes, and it leads towards inflated accuracy but poor recall rates for high-risk events [53]. The existing literature indicated that such models that are trained on imbalanced datasets faced a huge reduction in minority class detection rates by up to 18 to 25% unless cost-sensitive or resampling learning techniques are applied [3]. Furthermore, a lot of datasets are simulated or synthetic, which limits generalizability. However, the absence of a real-world and standardized cyber risk dataset can complicate benchmarking [11].

7.2. Model Interpretability and Trust Issues

Some advanced AI models, like deep learning and transformer-based architectures, often work as black boxes, and these models have achieved a high predictive accuracy that may frequently exceed 94%, but their decision-making processes remain opaque [3]. However, in high-stakes environments like critical infrastructure, healthcare, and finance, there is a lack of interpretability that decreases trust of stakeholders and limits regulatory acceptance [56]. The required empirical findings showed that security analysts are 30 to 40% more likely to rely on AI-generated risk scores when combined with explain ability tools like LIME or SHAP visualization [19]. In case of the absence of an appropriate transparent reasoning mechanism, it poses a massive threat to the region or under-utilization of risks in the operational environments.

7.3. Scalability and Real-Time Deployment Challenges

Under controlled experimental environments, AI models performed well. But when applied on the enterprise level, they are challenged by enormous difficulties [8]. It is because huge organizations are processing more than a million security events at a time. Therefore, it requires high computational power and capacity with low-latency processing. Some real-time deployment constraints indicate that it is challenging to include continuous model retention requirements, as integration complexity increases due to high infrastructure costs and legacy systems [3]. The research mentioned that model performance is decreased by 5 to 12% when the company transformed from online testing to a real-time prediction environment, and it is due to data drift and operational variability [53]. However, scalability remains a significant practical limitation.

7.4 Adversarial Attacks against AI Systems

The problem is that these AI systems are the main targets of cyber-attacks. The reason behind it is that adversarial machine learning techniques can manipulate the input data to deceive AI models. If there is a slight misinterpretation present in the network traffic, then the detection accuracy rate by up to 20% in some deep learning models [8]. Secondly, poisoning attacks are links with adversaries inject malicious samples into training data and it can enhance threats to model reliability.

To handle this problem, various defense strategies like adversarial training and robust optimization are considered emerging but they are highly expensive and not properly explored in risk-specific scenarios.

7.5. Ethical, Regulatory, and Legal Concerns

Such Risk Assessment systems that are powered by AI-driven technologies raise a lot of concerns regarding accountability, bias, and privacy. Moreover, automated risk scoring is highly powerful against certain groups and other organizational units when training data is based on historical biases [21]. Moreover, there is a need for high compliance, transparency, and accountability for handling regulatory frameworks. Hence, organizations must ensure that AI systems comply positively with cybersecurity standards and regulations.

8. Future Research Directions

From the last decade these AI-assisted cyber risk management systems are highly powerful and they showed some remarkable results, but still some research gaps are present that offer new opportunities for future exploration.

8.1. Requirement for Standardized Risk Assessment Framework

Latest studies employ heterogeneous methodologies, evaluation metrics and datasets that limit reproducibility. Overall, there is no universally accepted and perfect AI-driven cyber-risk assessment framework that can integrate threat intelligence, asset valuation, vulnerability metrics, and financial impact estimation [53]. Therefore, the required future research needs to focus on developing modular and standardized frameworks capable of integrating with established governance models like FAIR and NIST by incorporating AI-driven probabilistic modelling.

8.2. Explainable and Transparent AI Models

In the past few years, explainable AI techniques have gained a lot of attention, but most implementations remain superficial. Hence, research is required to develop inherently interpretable risk models rather than only depend on post-hoc explainable tools [11]. For future studies, hybrid architectures are required that combine rule-based logic with machine learning so that they can provide a perfect balance between transparency and performance. By developing a quantifiable trust metrics for AI-based risk models is considered another vital future direction.

8.3. Federated and Privacy-Preserving Learning

The data taken from cyber-risk platforms is highly sensitive and difficult to share across organizations. However, federated learning is highly useful because it offers a promising approach by enabling decentralized model training without exposing raw data [21]. Based on preliminary research, it can be seen that federated models can achieve 90 to 95% of centralized model accuracy by maintaining high data privacy [6]. Secondly, some problems present related to communication overhead, convergence, and heterogeneity are not solved properly.

9. Discussion

This conceptual study explained that AI-techniques like deep learning, ensemble learning, and hybrid architectures had achieved high predictive performance in handling cyber risk assessment tasks with reported accuracies ranging from 88 to 97%. New risk models are constantly more precise and scalable as compared to the outdated qualitative models. The study also contributes to the theoretical knowledge, as it synthesizes AI methods in the framework of a single cyber risk management model. The article identifies some of the important changes in a fixed risk scoring to predictive, adaptive, and autonomous systems. Second, it differentiates new intersections, which can be found in the space of federated learning, explainable AI, and generative modelling of cybersecurity. The AI-based risk management has a measurable benefit to the industries. These are improved prioritization, shorter detection and optimum allocation of resources. However, to reach a large-scale adoption, organizations have to contend with a variety of deployment challenges, present potent governance frameworks, and render them understandable. Governance considers that the policy makers should stimulate the standardized assessment systems, international cooperation and demand of transparency. Hence, regulatory alignment will be very critical in order to make AI deployment in cybersecurity ethical and secure.

10. Conclusion

It is concluded that AI-Assisted cyber risk assessment and management is considered the best technology to handle cyber security risk. The large enterprises can easily increase their cyber-security and reactive defence to more predictive and proactive risk control by integrating various technologies such as machine learning, deep learning, and adaptive decision models. Overall, the study investigated the current trends of 2022-2026 in AI-driven methods of cyber risk, and it presents quantitative comparisons between different methods and describes some significant limitations. This research provides an elaborate source of reference to practitioners and researchers through the synthesis of empirical results. The AI technology is an important improvement in risk management. However, it can be effective only when it comes to interpretability, good data, good defences and alignment of governance to adversarial manipulation. Subsequently, the additional research of the interdisciplinary one founded on the combination of risk theory and computer science and regulatory policy will be of significant value in the future to make the maximum of AI in the security of modern digital

Declarations

Acknowledgements

The work was supported by King Faisal University, Saudi Arabia. The authors would like to thank the scholastic reviewers for their intuitive comments and sagacious suggestions to improve the quality and clarity of this paper.

Funding
NA

Contributions

D.A; K.A; H.R; Conceptualization, D.A; K.A; H.R; Investigation, D.A; K.A; Writing (Original Draft), H.R; Writing (Review and Editing) & Supervision, H.R; Project Administration.

Ethics declarations

This article does not contain any studies with human participants or animals performed by any of the authors.

Data Availability Statement

The authors thank their co-authors for guidance, peer reviewers for valuable feedback, and all participants for their contributions. They also acknowledge the support of the Lincoln University College Doctoral Program and appreciate the editorial team and anonymous reviewers of the Journal of Security Risk Management for their constructive input.

Use of Generative Artificial Intelligence

During the preparation of this work the authors used AI tools in order to assist in language refinement, proofreading, and structural suggestions during the preparation of this manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Competing interests

All authors declare no competing interests.

References

[1] MANDYCH, O., ZAIKA, S., ZAIKA, O., ZHYLIAKOVA, O., & BLYZNIUK, Oksana. (2023). RISK MANAGEMENT OF INNOVATION ACTIVITIES IN THE DIGITAL ECOSYSTEM. Innovations in the Scientific, Technical and Social Ecosystems, 1(6), 24–45. https://doi.org/10.56378/mozs20231805

[2] Huaman, N., von Skarczinski, B., Stransky, C., Wermke, D., Acar, Y., Dreißigacker, A., & Fahl, S. (2021). A large-scale interview study on information security in and attacks against small and medium-sized enterprises. In Proceedings

of the 30th USENIX Security Symposium (USENIX Security 21) (pp. 1235–1252). USENIX Association. https://www.usenix.org/conference/usenixsecurity21/presentation/huaman

[3] Cobos, E. V., Cakir, S., Straub, S., Qiang, C. Z., & Torgusson, C. (2024). A review of the economic costs of cyber incidents. *World Bank, Washington, DC, USA, Rep*, *193919*.

[4] Hanisch, R., Kaiser, D., Yuan, A., Medina-Smith, A., Carroll, B., & Campo, E. (2024). NIST Research Data Framework (RDaF) (, Ed.). National Institute of Standards and Technology. https://doi.org/10.6028/nist.sp.1500-18r2

[5] Nechita, R.-M., Stochioiu, F.-P.-G., & Grecu, I. (2025). Identifying and Prioritising Factors for Effective Decision-Making in Data-Driven Organisations: A DEMATEL Approach. Systems, 13(8), 687. https://doi.org/10.3390/systems13080687

[6] Beckley, J. (2025). Advanced Risk Assessment Techniques: Merging Data-Driven Analytics with Expert Insights to Navigate Uncertain Decision-Making Processes. International Journal of Research Publication and Reviews, 6(3), 1454–1471. https://doi.org/10.55248/gengpi.6.0325.1148

[7] Kalla, D., Samaah, F., & Kuraku, Dr. S. (2021). Enhancing cyber security by predicting malwares using supervised machine learning models. International Journal of Computing and Artificial Intelligence, 2(2), 55–62. https://doi.org/10.33545/27076571.2021.v2.i2a.71

[8] Tan, Z., Nagar, U. T., He, X., Nanda, P., Liu, R. P., Wang, S., & Hu, J. (2014). Enhancing Big Data Security with Collaborative Intrusion Detection. IEEE Cloud Computing, 1(3), 27–33. https://doi.org/10.1109/mcc.2014.53

[9] Mohammad, A. S., & Pradhan, M. R. (2021). Machine learning with big data analytics for cloud security. Computers and Electrical Engineering, 96, 107527. https://doi.org/10.1016/j.compeleceng.2021.107527

[10] Sood, S. K. (2012). A combined approach to ensure data security in cloud computing. Journal of Network and Computer Applications, 35(6), 1831–1838. https://doi.org/10.1016/j.jnca.2012.07.007

[11] Kumar, P. R., Raj, P. H., & Jelciana, P. (2018). Exploring Data Security Issues and Solutions in Cloud Computing. Procedia Computer Science, 125, 691–697. https://doi.org/10.1016/j.procs.2017.12.089

[12] Alkasassbeh, M., Omoush, E. H., Almseidin, M., & Aldweesh, A. (2025). A Self-Adaptive Intrusion Detection System for Zero-Day Attacks Using Deep Q-Networks. IEEE Access, 13, 174280–174296. https://doi.org/10.1109/access.2025.3617792

[13] Chadwick, D. W., Fan, W., Costantino, G., de Lemos, R., Di Cerbo, F., Herwono, I., Manea, M., Mori, P., Sajjad, A., & Wang, X.-S. (2020). A cloud-edge based data security architecture for sharing and analysing cyber threat information. Future Generation Computer Systems, 102, 710–722. https://doi.org/10.1016/j.future.2019.06.026

[14] Stergiou, C., Psannis, K. E., Gupta, B. B., & Ishibashi, Y. (2018). Security, privacy & efficiency of sustainable Cloud Computing for Big Data & IoT. Sustainable Computing: Informatics and Systems, 19, 174–184. https://doi.org/10.1016/j.suscom.2018.06.003

[15] Bahani, K., Moujabbir, M., & Ramdani, M. (2021). An accurate fuzzy rule-based classification systems for heart disease diagnosis. Scientific African, 14, e01019. https://doi.org/10.1016/j.sciaf.2021.e01019

[16] Wei, X., & Dong, Y. (2025). A hybrid approach combining Bayesian networks and logistic regression for enhancing risk assessment. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-10291-9

[17] Li, J., Li, B., Wo, T., Hu, C., Huai, J., Liu, L., & Lam, K. P. (2012). CyberGuarder: A virtualization security assurance architecture for green cloud computing. Future Generation Computer Systems, 28(2), 379–390. https://doi.org/10.1016/j.future.2011.04.012

[18] Wang, R. (2017). Research on Data Security Technology Based on Cloud Storage. Procedia Engineering, 174, 1340–1355. https://doi.org/10.1016/j.proeng.2017.01.286

[19] Velmurugadass, P., Dhanasekaran, S., Shasi Anand, S., & Vasudevan, V. (2021). Enhancing Blockchain security in cloud computing with IoT environment using ECIES and cryptography hash algorithm. Materials Today: Proceedings, 37, 2653–2659. https://doi.org/10.1016/j.matpr.2020.08.519

[20] Wang, Z., Wang, N., Su, X., & Ge, S. (2020). An empirical study on business analytics affordances enhancing the management of cloud computing data security. International Journal of Information Management, 50, 387–394. https://doi.org/10.1016/j.ijinfomgt.2019.09.002

[21] Du, Y., Chatterjee, S., Bhattacharya, A., Dutta, A., & Halappanavar, M. (2023). Role of reinforcement learning for risk-based robust control of cyber-physical energy systems. Risk Analysis, 43(11), 2280–2297. Portico. https://doi.org/10.1111/risa.14104

[22] Gamaleldin, W., Attayyib, O., Alnfiai, M. M., Alotaibi, F. A., & Ming, R. (2025). A hybrid model based on CNN-LSTM for assessing the risk of increasing claims in insurance companies. PeerJ Computer Science, 11, e2830. Portico. https://doi.org/10.7717/peerj-cs.2830