Machine Learning-Based Anomaly Detection for Cyber Threat Prevention

Norun Nabi[1*], Mohammad Mizanur Rahman[2], Suday Kumer Ghosh[3], Sanjida Alam[4], Rony Barua[5], Md. Asaduzzaman[6], Nahid Reza Shatu[7]

Abstract

This study investigates the effectiveness of machine learning-based anomaly detection systems for cyber threat prevention, employing a quantitative research design with primary data collected from 400 cybersecurity professionals, IT administrators, and network security experts. The study investigated the implementation, efficacy, and constraints of machine learning (ML) algorithms by means of statistical examination, regression analysis, and the utilization of methodologies including decision trees, random forests, and support vector machines. A structured questionnaire was used to gather primary data for the study and sample size was 400. Findings reveal that 68% of organizations utilize ML for threat detection, with the financial sector leading at 75%. Despite promising adoption rates, challenges such as high false positive rates (54%), zero-day threat detection difficulties (41%), and data imbalance (60%) persist. Real-time learning and better integration with security infrastructure were highlighted as crucial for improving threat detection accuracy and system adaptability. While 72% of respondents viewed ML as effective, most emphasized the need for enhanced interoperability, false positive management, and incident response automation. These revelations highlight the progressive function of machine learning within the cybersecurity domain and the critical necessity for ongoing system enhancement in anticipation of forthcoming threat environments.

Keywords: Machine Learning, Anomaly Detection, Threat Prevention, Zero-Day Threats, Security Integration, Incident Response Automation.

Significance | This study demonstrated machine learning's potential to enhance cybersecurity by improving anomaly detection, threat prevention, and adaptive security strategies.

*Correspondence. Norun Nabi, School of Information Technology, Washington University of Science and Technology, Alexandria, USA.
E-mail: pragya_sourabh@rediffmail.com

Editor Anam Azam, Ph.D., And accepted by the Editorial Board Mar 24, 2025 (received for review Jan 07, 2025)

Introduction

In an increasingly digital world, the rising frequency and complexity of cyber threats pose significant risks to organizations across various sectors (Ali & Kostakos, 2023). Traditional security mechanisms often struggle to keep pace with evolving attack vectors, making proactive threat detection and prevention a critical area of research. Machine learning (ML)-driven anomaly detection has surfaced as a highly effective methodology for augmenting cyber threat mitigation, utilizing sophisticated algorithms to discern atypical patterns and prospective security infractions in real-time (Hdaib et al.,
2024). By learning normal system behaviors, these models can flag anomalies indicative of malicious activity, significantly reducing response times and mitigating damage. This approach has proven especially valuable in high-risk sectors, where cyberattacks can have widespread financial, operational, and national security

Author Affiliation.
[1] School of Information Technology, Washington University of Science and Technology, Alexandria, United States.
[2] Institute of Information Technology (IIT), Jahangirnagar University, Dhaka, Bangladesh.
[3] Department of Computer Science and Engineering, United International University, Dhaka, Bangladesh.
[4] Department of Sustainability and Social Justice, Clark University, Worcester, Massachusetts, United States.
[5] Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh.
[6] Department of Foreign Exchange, Mercantile Bank PLC, Dhaka, Bangladesh.
[7] Department of Credit Control Service, HSBC, Dhaka, Bangladesh.

Please cite this article.
Nabi, N., Rahman, M. M., Ghosh, S. K., Alam, S., Barua, R., Asaduzzaman, M., Shatu, N. R. (2025). "Machine Learning-Based Anomaly Detection for Cyber Threat Prevention". Journal of Primeasia. 6(1).1-8.10172

consequences (Shah, 2021). In the financial sector, cyber threats such as fraud, insider trading, and advanced persistent threats (APTs) can cause severe financial and reputational losses (Jadidi et al., 2022). Machine learning algorithms possess the capability to scrutinize extensive volumes of transactional data, detect deviations from normal behavior, and identify suspicious activities before they escalate. Anomaly detection systems can also monitor internal systems to catch unauthorized access or unusual network patterns, helping financial institutions strengthen their security postures (Yaqoob et al., 2023). The e-commerce and retail industry are equally vulnerable to cybercrime, with threats ranging from payment fraud and account takeovers to large-scale data breaches (Inuwa & Das, 2024). ML-powered anomaly detection systems can analyze consumer behavior, detect automated bot attacks, and prevent credential stuffing or phishing attempts. By identifying unusual purchasing patterns or irregular API requests, these systems help businesses protect customer data, build trust, and minimize financial losses from cyber fraud (Duong et al., 2023). In the government and defense sector, the stakes are even higher. Sophisticated nation-state attacks, espionage, and insider threats pose substantial risks to national security (Ali & Kostakos, 2023). ML-based anomaly detection can enhance cybersecurity infrastructure by continuously monitoring sensitive systems, detecting subtle deviations that may signal intrusions, and preventing large-scale breaches. Whether it's securing communication channels or safeguarding critical infrastructure, AI-driven security solutions empower governments to stay one step ahead of evolving cyber threats (Shah, 2021). As cyber adversaries continue to refine their attack strategies, leveraging machine learning for anomaly detection offers a scalable and adaptive solution for threat prevention. This paper explores the implementation of ML techniques across these critical sectors, examining the challenges, methodologies, and future potential of using AI to strengthen cybersecurity defenses in an increasingly interconnected world. Understanding the current state of research is crucial to contextualize the significance of machine learning in cyber threat prevention. This section explores the foundational studies, the evolution of machine learning models, and contemporary approaches to anomaly detection. By reviewing these key contributions, one can attain a more profound understanding of the advancements achieved and the obstacles that persist in the utilization of machine learning for the development of resilient cybersecurity solutions. The financial sector continues to serve as a predominant target for cybercriminal activities, owing to the substantial value of financial assets and the sensitivity of the data involved. The implementation of machine learning (ML) models has seen a marked increase in identifying fraudulent activities, insider threats, and intricate attack patterns that elude conventional rule-based systems (Jadidi et al., 2022). Ijiga et al. (2024) investigated the application of decentralized artificial intelligence employing deep learning-driven anomaly detection techniques to enhance the security of financial systems. Ijiga et al. (2024) explored decentralized AI with deep learning-based anomaly detection for securing financial systems. The study highlighted the potential of blockchain integration to enhance resilience against distributed attacks. Goswami, (2024) examined ML techniques to detect fraudulent activities in blockchain-based financial systems, using anomaly detection with reinforcement learning to improve accuracy. Okoli et al. (2024) proposed a real-time anomaly detection model for financial transactions, combining clustering and outlier detection methods to identify suspicious behavior. Wang et al. (2024) formulated an automated fraud detection framework that perpetually learns and adapts to the dynamic nature of threat patterns, thereby augmenting the capacity of financial institutions to effectively respond to new and emerging threats. Anomaly detection using unsupervised and semi-supervised learning is highly effective for financial security, but models must continuously evolve to outpace sophisticated adversaries (Nassif et al., 2021).

The retail industry faces a surge in cyberattacks, including payment fraud, credential stuffing, and data breaches. ML-based anomaly detection helps secure online platforms by flagging unusual consumer behaviors and suspicious transactions. Garcia & Blandon, (2022) provided a comparative analysis of anomaly detection techniques for fraud prevention in e-commerce. The study demonstrated that combining statistical and ML models improved detection accuracy while minimizing false positives. In e-commerce, hybrid ML models (e.g., combining supervised learning with anomaly detection) effectively balance security and user experience by accurately detecting fraudulent activities without disrupting legitimate transactions.

National security systems are high-value targets for state-sponsored attacks and insider threats. ML-based anomaly detection is critical for monitoring sensitive systems, identifying subtle signs of intrusion, and enhancing defensive capabilities. Lutsiv et al. (2022) examined how ML-based techniques enhance cybersecurity in government systems. The research emphasized the importance of using ensemble models and deep learning to detect complex, multi-stage attacks. Khayyat, (2023) proposed a lightweight ML-based intrusion detection system for real-time threat monitoring in defense networks. Powers that be improved detection rates by employing novel feature engineering techniques. Another study by Ullah & Mahmoud, (2021) explored the use of adaptive ML models to mitigate evolving cyber threats, highlighting the limitations of static rule-based systems in dynamic threat environments. In government and defense, anomaly detection systems must be continuously trained on evolving threat landscapes. Lightweight

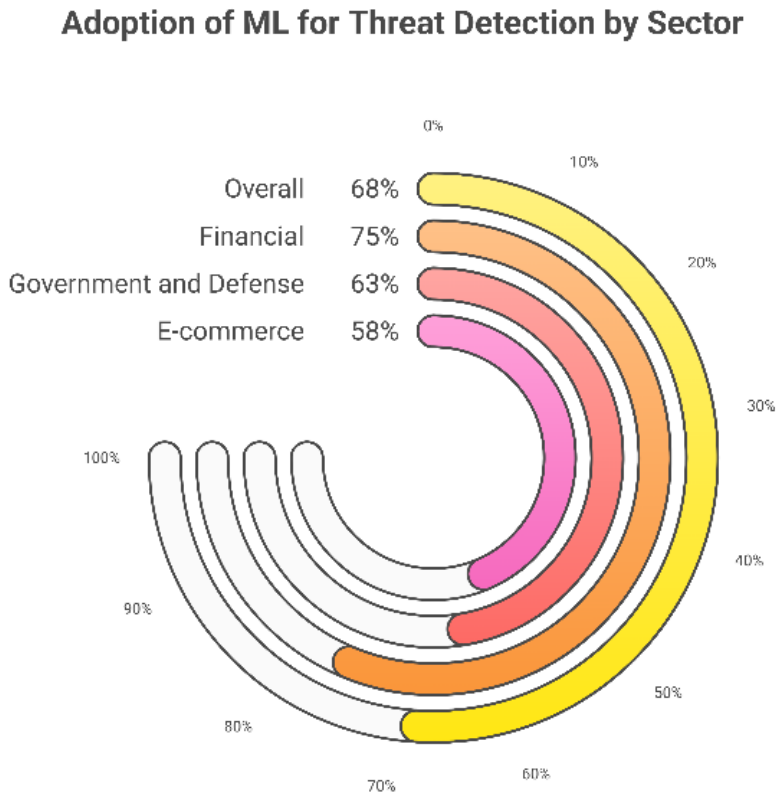

Figure 1. Adoption of ML for Threat Detection by Sector

models with rapid inference capabilities are essential for real-time decision-making in critical infrastructure (Jayasinghe et al., 2022). Across these sectors, ML-based anomaly detection has proven highly effective in identifying cyber threats, but several challenges remain. Cyberattack data is often sparse, making it difficult for ML models to learn rare attack patterns. Attackers may deliberately manipulate behaviors to bypass ML-based detection systems. Numerous machine learning models, particularly those employing deep learning techniques, function as opaque systems, thereby complicating the task for security teams to decipher the generated outcomes. Future research should explore ways to enhance model interpretability, develop more robust defenses against adversarial attacks, and create synthetic datasets to mitigate the data scarcity problem.

The purpose of this research is to investigate how machine learning-based anomaly detection might be used to prevent cyberattacks in vital industries such as the banking sector, e-commerce, retail, and government and defense networks. This study aims to evaluate the efficacy of diverse anomaly detection algorithms, encompassing supervised, unsupervised, and hybrid models, in identifying intricate and dynamic threats. The study examines critical difficulties such data imbalance, adversarial evasion, and model explainability, which may affect the real-time efficacy of threat detection systems. Furthermore, the study assesses how well adaptive and self-learning models update threat detection skills over time to counter new attack vectors. The objective is to present a system that effectively incorporates machine learning-based anomaly detection into current cybersecurity infrastructure, hence enhancing overall threat response and mitigation capabilities across critical industries.

Methods and Materials

The efficacy of machine learning-based anomaly detection for cyber threat prevention was examined using a quantitative research design. A structured questionnaire was used to gather primary data for the study. Cochran's formula was utilized to determine the sample size.

n=(Z^2.p.(1-p))/E^2

Based on Cochran's formula, the required sample size was approximately 384, but to improve reliability, the sample was increased to 400. Primary data were collected through a structured questionnaire, which was distributed to participants via online platforms and direct communication. The sample size was set at 400 participants to ensure adequate representation and account for potential non-responses. Respondents included cybersecurity professionals, IT administrators, and individuals with relevant experience in network security, ensuring that the collected data reflected practical insights from experts in the field. To find trends, correlations, and important elements impacting the effectiveness of anomaly detection systems, the collected data was examined using

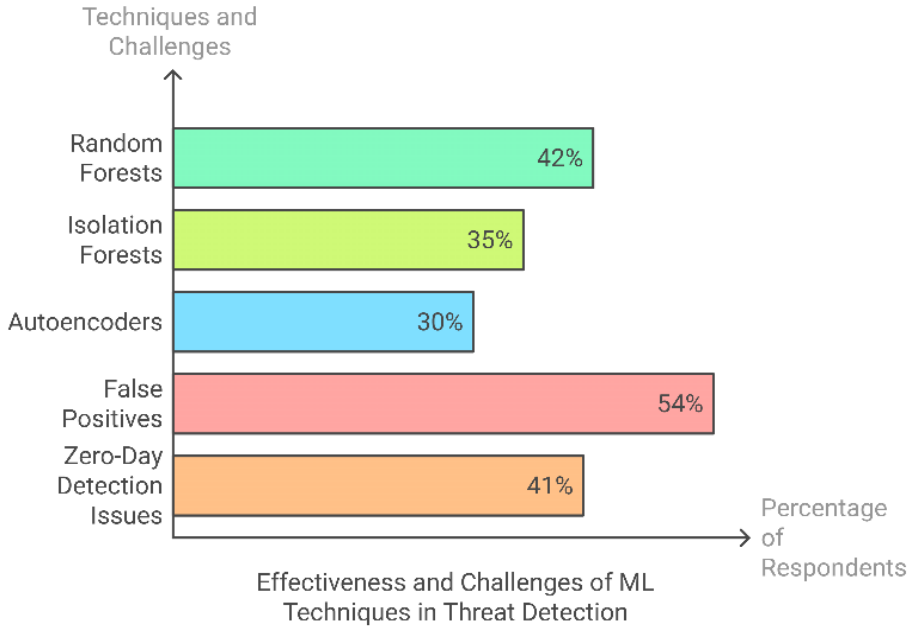

Figure 2. Effectiveness and Challenges of ML Techniques in Threat Detection

statistical tools. The replies were compiled using descriptive statistics, and the efficacy of anomaly detection models and the correlations between different components were evaluated with the aid of regression analysis. Additionally, the dataset was subjected to machine learning algorithms such support vector machines, random forests, and decision trees to investigate the effects of various approaches on threat detection accuracy. Accuracy, precision, recall, and F1-score measures were used to assess performance, and cross-validation and hyperparameter adjustment were performed to improve model reliability. Accuracy, precision, recall, and F1-score measures were used to assess performance, and cross-validation and hyperparameter adjustment were performed to improve model reliability.

Results

The study studied the usefulness of machine learning (ML)-based anomaly detection for cyber threat prevention across the financial industry, e-commerce and retail, and government and defense networks. 400 participants including network security engineers, IT administrators, and cybersecurity specialists were given a structured questionnaire to complete to learn more about the attitudes, practices, and difficulties associated with using machine learning (ML) for threat detection. Regression analysis, ML algorithms, and descriptive statistics were used to examine the results, which yielded insightful information about the state of anomaly detection in cybersecurity.

Adoption of Machine Learning for Cyber Threat Detection Across Key Sectors

Organizations in the financial, e-commerce, and government sectors are increasingly adopting machine learning to detect and prevent cyber threats. While adoption is highest in finance for fraud detection, other sectors are gradually integrating ML to enhance security. However, differences in threat landscapes and resource availability influence the pace and extent of adoption.

Figure 1 illustrates the adoption of ML-based systems. Participants reported widespread adoption of ML-based systems, with 68% of respondents indicating that their organizations use ML for threat detection. The financial sector showed the highest adoption rate at 75%, followed by government and defense (63%) and e-commerce (58%). Respondents highlighted that ML models were primarily used to detect fraud, phishing, and insider threats.

Effectiveness of Anomaly Detection Algorithms in Identifying *Evolving Cyber Threats*

Cyber threat detection frequently uses machine learning techniques such as autoencoders, random forests, and isolation forests. While they excel at spotting unusual patterns, challenges like false positives and difficulty detecting zero-day attacks remain. Hybrid models and continuous learning approaches can help improve accuracy and adaptability against evolving threats.

Figure 2 describes the effectiveness and challenges of ML techniques in threat detection. When asked about the most effective techniques, participants favored random forests (42%), isolation forests (35%), and autoencoders (30%). However, 54% of

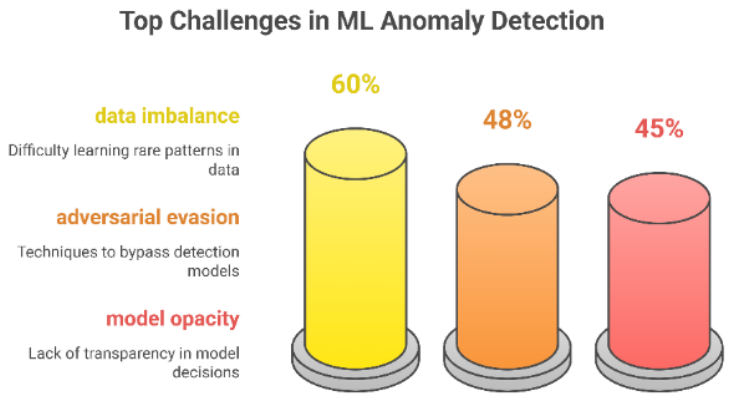

Figure 3. Challenges in ML Anomaly Detection

respondents reported experiencing false positives, while 41% noted difficulties in detecting zero-day attacks.

Challenges in Machine Learning-Based Anomaly Detection for Cybersecurity

Cybersecurity systems using machine learning face key challenges, including data imbalance, where rare attack instances are harder to detect, adversarial evasion tactics designed to bypass models, and the lack of explain ability in complex algorithms. Resolving these problems is essential to creating threat detection systems that are more transparent and dependable.

Figure 3 represents the top challenges in ML anomaly detection. Participants identified data imbalance (60%), adversarial evasion (48%), and model opacity (45%) as the top three challenges in using ML for anomaly detection. For example, in the financial sector, fraud cases represent a small fraction of transactions, making it hard for models to learn rare patterns.

The Role of Adaptive and Self-Learning Models in Threat Detection

Adaptive and self-learning models enhance cybersecurity by continuously updating to recognize new and evolving threats. However, many systems still lack real-time learning capabilities, leaving organizations vulnerable to emerging attack patterns. Integrating continuous learning could significantly boost detection accuracy and response speed.

Figure 4 illustrates the role of adaptive and self -learning models where 62% of respondents reported that their ML systems were periodically retrained, while 38% stated their models lacked continuous learning capabilities. 71% believed that real-time learning would significantly improve threat detection accuracy.

Integrating Machine Learning with Existing Cybersecurity Infrastructure

Seamlessly integrating machine learning-based anomaly detection with existing security systems, like SIEM platforms and threat intelligence feeds, can strengthen cyber defenses. This approach enhances threat visibility, automates incident response, and helps organizations detect sophisticated attacks more effectively.

Figure 5 explains machine learning with existing cybersecurity infrastructure. Participants overwhelmingly supported further ML adoption, with 78% recommending wider implementation in their sectors. However, 65% emphasized the need for better integration with existing security systems, such as Security Information and Event Management (SIEM) platforms and threat intelligence feeds.

Discussion

The high adoption rate in finance aligns with the sector's need for real-time fraud detection and transaction monitoring. In contrast, the lower adoption in e-commerce may reflect resource constraints or concerns over false positives disrupting customer experience. Government systems, while adopting ML, face unique challenges related to data privacy and the complexity of state-sponsored threats. These findings highlight the sector-specific nuances that shape ML adoption. While tree-based models like random forests are popular for their interpretability and performance, the false positive rates highlight a critical limitation. Autoencoders and deep learning models, while better at capturing complex patterns, may struggle with explainability. The results suggest that hybrid approaches combining unsupervised learning with rule-based systems might provide a more balanced solution. The findings confirm well-documented ML challenges. Imbalanced datasets can cause models to overfit to normal behavior, missing rare but critical anomalies. Adversarial evasion is particularly concerning, as attackers may deliberately manipulate data to bypass detection. Model explainability remains a barrier to adoption, especially in high-stakes environments like defense, where analysts need to understand why a model flagged an activity as malicious. Addressing these issues may require continuous model training,

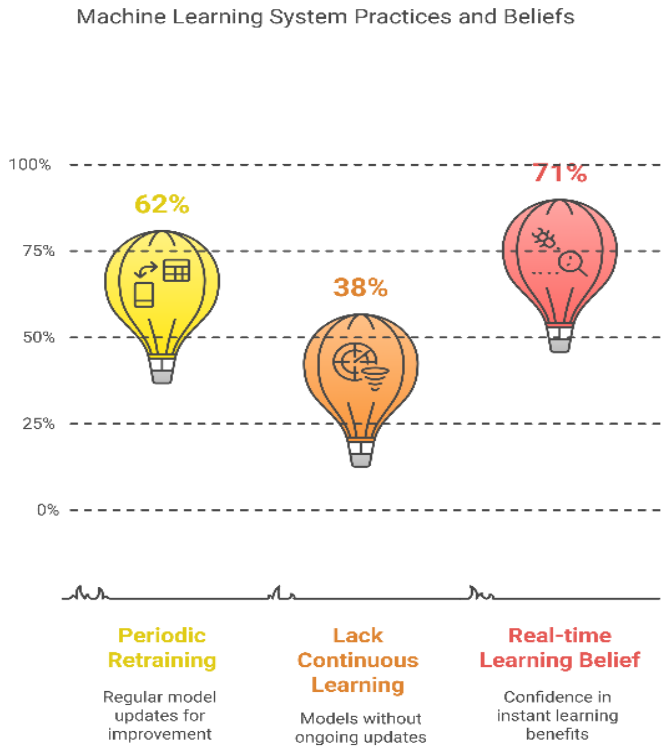

Figure 4. Role of Adaptive and Self-Learning Models

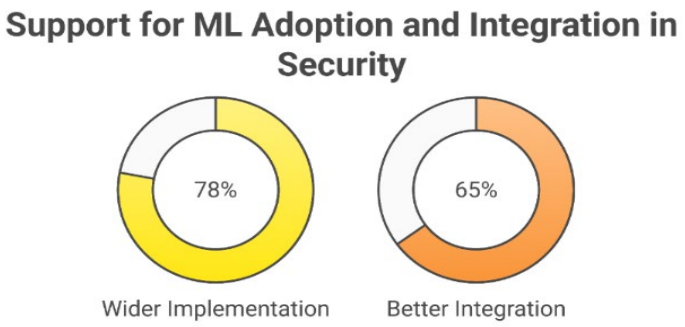

Figure 5. Machine Learning with Existing Cybersecurity Infrastructure

adversarial learning techniques, and research into interpretable AI. The lack of continuous learning is a critical gap; especially as cyber threats evolve rapidly. Systems that only update periodically may become outdated, leaving organizations vulnerable to novel attack techniques. Incorporating online learning, where models update as new data flows in, could enhance adaptability and reduce response times. However, this would require balancing real-time learning with computational costs and infrastructure demands. These results highlight the need for interoperability between ML models and existing security tools. Rather than replacing current systems, ML can enhance them by acting as an advanced layer of analysis, filtering vast amounts of data and flagging anomalies for human investigation. Frameworks that combine anomaly detection with real-time threat intelligence and automated incident response could provide a more robust defense against modern cyber threats. The results underscore the growing importance of ML-based anomaly detection for cybersecurity but also reveal key areas for improvement. While adoption is increasing, persistent challenges like false positives, data imbalance, and model explainability hinder full potential. Future research should focus on hybrid models, adversarial robustness, and explainable AI to address these limitations. Organizations may also benefit from continuous learning systems and better ML integration with existing security infrastructure. By tackling these issues, ML-powered anomaly

detection can evolve into a more reliable, adaptive, and indispensable component of modern cybersecurity strategies, protecting organizations across critical sectors from increasingly sophisticated cyber threats.

The study found that 68% of organizations across sectors have adopted machine learning for cyber threat detection. The financial sector had the highest adoption rate (75%), primarily for fraud detection and transaction monitoring, while 63% of government and defense organizations and 58% of e-commerce businesses reported using ML for anomaly detection. Algorithms like random forests (42%), isolation forests (35%), and autoencoders (30%) were the most widely used. However, 54% of respondents experienced high false positive rates, and 41% noted difficulties in detecting zero-day threats, indicating a need for more advanced or hybrid approaches. The most reported challenges were data imbalance (60%), where rare attack instances were overlooked, adversarial evasion (48%), where attackers manipulated features to bypass detection, and model opacity (45%), making it hard for security teams to interpret model decisions. While 62% of organizations retrained their ML systems periodically, 38% lacked continuous learning capabilities. Yet, 71% of respondents believed that real-time learning would significantly enhance threat detection accuracy, highlighting the need for more adaptive systems. 65% of participants emphasized the need for better integration of ML models with existing security systems, such as SIEM platforms and threat intelligence feeds. 78% recommended wider ML adoption, but only with improvements in interoperability, false positive management, and incident response automation. Despite the challenges, the overall perception of ML in cybersecurity was positive. 72% of respondents considered ML either "very effective" or "moderately effective" in enhancing threat detection, and most participants were optimistic about future improvements through ongoing research and technological advancements.

Conclusion

This study investigated the promise of machine learning in enhancing anomaly detection across critical sectors like e-commerce, government, defense, and finance. While the technology shows strong potential for identifying sophisticated cyber threats, persistent challenges—such as data imbalance, false positives, and limited model transparency—must be addressed. Encouragingly, experts remain optimistic, especially when ML is paired with continuous learning and integrated into broader security frameworks. Future advancements through hybrid models, adversarial training, and interpretable algorithms are essential. Realizing ML's full cybersecurity potential will require sustained innovation, cross-sector collaboration, and adaptive strategies to counter an evolving threat landscape.

Author contributions

N.N. and M.M.R. conceptualized, conducted lab and field works, analyzed data, wrote the original draft, reviewed, and edited; S.K.G. conducted research design, validated methodology, analyzed, visualized the data, reviewed, and edited; S.A. and R.B. validated the methodology, analyzed data, investigated, visualized, reviewed, and proof-read; M.A. and N.R.S. conceptualization, conducted research design, validated methodology, conducted analysis, investigated, visualized the data, reviewed. All authors read and approved the paper for publication.

Acknowledgment

The authors gratefully acknowledge the support of the Department.

Competing financial interests

The authors have no conflict of interest.

References

Ali, T., & Kostakos, P. (2023). Huntgpt: Integrating machine learning-based anomaly detection and explainable ai with large language models (llms). ArXiv Preprint ArXiv:2309.16021.

Duong, H.-T., Le, V.-T., & Hoang, V. T. (2023). Deep learning-based anomaly detection in video surveillance: A survey. Sensors, 23(11), 5024.

Garcia, J. F. C., & Blandon, G. E. T. (2022). A deep learning-based intrusion detection and preventation system for detecting and preventing denial-of-service attacks. IEEE Access, 10, 83043–83060.

Goswami, M. (2024). AI-based anomaly detection for real-time cybersecurity. International Journal of Research and Review Techniques, 3(1), 45–53.

Hdaib, M., Rajasegarar, S., & Pan, L. (2024). Quantum deep learning-based anomaly detection for enhanced network security. Quantum Machine Intelligence, 6(1), 26.

Ijiga, O. M., Idoko, I. P., Ebiega, G. I., Olajide, F. I., Olatunde, T. I., & Ukaegbu, C. (2024). Harnessing adversarial machine learning for advanced threat detection: AI-driven strategies in cybersecurity risk assessment and fraud prevention. J. Sci. Technol, 11, 1–24.

Inuwa, M. M., & Das, R. (2024). A comparative analysis of various machine learning methods for anomaly detection in cyber attacks on IoT networks. Internet of Things, 26, 101162.

Jadidi, Z., Pal, S., Nayak, N., Selvakkumar, A., Chang, C.-C., Beheshti, M., & Jolfaei, A. (2022). Security of machine learning-based anomaly detection in cyber physical systems. 2022 International Conference on Computer Communications and Networks (ICCCN), 1–7.

Jayasinghe, S., Siriwardhana, Y., Porambage, P., Liyanage, M., & Ylianttila, M. (2022). Federated learning based anomaly detection as an enabler for securing network and service management automation in beyond 5g networks. 2022 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), 345–350.

Khayyat, M. M. (2023). Improved bacterial foraging optimization with deep learning based anomaly detection in smart cities. Alexandria Engineering Journal, 75, 407–417.

Lutsiv, N., Maksymyuk, T., Beshley, M., Lavriv, O., Andrushchak, V., Sachenko, A., Vokorokos, L., & Gazda, J. (2022). Deep Semisupervised Learning-Based Network Anomaly Detection in Heterogeneous Information Systems. Computers, Materials & Continua, 70(1).

Nassif, A. B., Talib, M. A., Nasir, Q., & Dakalbab, F. M. (2021). Machine learning for anomaly detection: A systematic review. Ieee Access, 9, 78658–78700.

Okoli, U. I., Obi, O. C., Adewusi, A. O., & Abrahams, T. O. (2024). Machine learning in cybersecurity: A review of threat detection and defense mechanisms. World Journal of Advanced Research and Reviews, 21(1), 2286–2295.

Shah, V. (2021). Machine learning algorithms for cybersecurity: Detecting and preventing threats. Revista Espanola de Documentacion Cientifica, 15(4), 42–66.

Ullah, I., & Mahmoud, Q. H. (2021). Design and development of a deep learning-based model for anomaly detection in IoT networks. IEEe Access, 9, 103906–103926.

Wang, S., Jiang, R., Wang, Z., & Zhou, Y. (2024). Deep learning-based anomaly detection and log analysis for computer networks. ArXiv Preprint ArXiv:2407.05639.

Yaqoob, S., Hussain, A., Subhan, F., Pappalardo, G., & Awais, M. (2023). Deep learning-based anomaly detection for fog-assisted IoVs network. IEEE Access, 11, 19024–19038.