International Journal of Computer Science Engineering Techniques – Volume 10 Issue 2, March-April - 2026

Predictive Machine Learning Models for Forecasting Exploitable Network Vulnerabilities

Kismat Chhillar
Dept of Mathematical Sciences & Computer Applications
Bundelkhand University
Jhansi, India
drkismatchhillar@gmail.com

Deepak Tomar
System Analyst, Computer Center
Bundelkhand University
Jhansi, India
dr.deepak@bujhansi.ac.in

Dhruv Srivastava
Dept of Mathematical Sciences & Computer Applications
Bundelkhand University
Jhansi, India
dhruv.srivastava77@gmail.com

Anil Kewat
Dept of Mathematical Sciences & Computer Applications
Bundelkhand University
Jhansi, India
anil.kewat2007@gmail.com

Abstract— **This paper explores the application of predictive machine learning models for identifying network vulnerabilities that could be exploited, with the goal of enhancing proactive cybersecurity measures. By leveraging historical vulnerability data, attack patterns, and network configurations, these models can forecast potential exploit scenarios before they manifest. This anticipatory approach substantially reduces risk exposure and enables organizations to fortify their security posture. The capability to forecast vulnerabilities accelerates patch prioritization and supports dynamic risk management that evolves with emerging threats. The study examines several advanced machine learning techniques, including ensemble approaches such as random forests and gradient boosting, as well as deep learning architectures like long short-term memory (LSTM) networks, which specialize in capturing temporal dependencies. The results demonstrate notable improvements in predictive accuracy and robustness, equipping cybersecurity professionals with data-driven insights into vulnerability evolution that guide more strategic resource allocation and effective threat mitigation.**

Keywords— Predictive models, Machine learning, Network vulnerabilities, Cybersecurity, Exploit forecasting, Intrusion detection

I. INTRODUCTION

The escalating complexity and frequency of cyberattacks have elevated network vulnerabilities to a critical concern for organizations worldwide. These vulnerabilities constitute weaknesses within hardware, software or configuration settings that adversaries can exploit to gain unauthorized access, disrupt operations, or compromise data integrity. Conventional vulnerability management approaches remain largely reactive, identifying and mitigating issues only after exploitation has occurred, which limits the overall effectiveness of defense strategies. As cyber threats grow more advanced, the need for predictive approaches capable of anticipating vulnerabilities before exploitation becomes increasingly urgent. Machine learning (ML) offers promising solutions by processing extensive and heterogeneous datasets to uncover patterns and correlations that human analysts may overlook. ML models extract features from historical vulnerability databases, exploit repositories, system configuration logs and real-time network traffic to enable predictive assessment. This proactive methodology allows cybersecurity teams to focus resources on vulnerabilities most likely to be exploited, thereby improving their responsiveness and operational efficiency.

Recent advances in ML, including ensemble learning techniques, deep neural networks and sequence-based models such as long short-term memory (LSTM) networks, have significantly enhanced predictive capabilities by modeling nonlinear dependencies and temporal dynamics. However, several challenges persist, such as addressing class imbalance where exploitable vulnerabilities form a minority within datasets, selecting meaningful predictive features, and adapting models to evolving attack patterns. Furthermore, integrating predictive systems into cybersecurity operations requires that models maintain interpretability to foster trust and support informed decision-making, while continuously updating with new data to remain relevant. Evaluating performance must also extend beyond accuracy metrics to encompass precision, recall and F-measures, ensuring the models' practical utility in real-world defense scenarios.

This paper offers a detailed look at predictive machine learning models aimed at forecasting network vulnerabilities that can be exploited. We dived into different modeling techniques, explored how data is sourced and processed, assessed experimental results from benchmark datasets, and considered the factors involved in deployment. In the next section, related works is discussed and our contributions are positioned within the wider research landscape. After that, the methodology that is used for data preparation and model development is outlined. Following this, experimental results are presented from our simulations and evaluations. Then, the implications and challenges are discussed that come with predictive modeling in the realm of cybersecurity. To wrap things up, our key findings are summarized and future research directions are suggested.

II. BACKGROUND AND RELATED WORK

Predictive modeling in cybersecurity has evolved considerably, progressing from traditional statistical and heuristic techniques to advanced machine learning methods designed to forecast exploitable vulnerabilities with greater precision [1] [2]. Early research primarily employed basic risk scoring systems and regression-based analyses to estimate both the severity of vulnerabilities and their likelihood of exploitation. These approaches, while useful,

often exhibited limitations in scalability and failed to capture the complex and dynamic characteristics of modern cyber threats. In response, more recent studies have adopted sophisticated machine learning algorithms such as support vector machines (SVM), random forests and logistic regression to analyze vulnerability datasets and predict exploitation potential [3] [4] [5]. Empirical investigations drawing on the National Vulnerability Database (NVD) and other publicly available datasets have demonstrated that machine learning models can successfully identify nuanced relationships among vulnerability attributes including attack vectors, impact metrics, and disclosure timelines [6] [7]. Furthermore, natural language processing (NLP) techniques have been integrated to evaluate textual data from vulnerability descriptions, exploit scripts and security advisories, thereby enriching feature representations and enhancing the predictive performance of these models.

Deep learning, particularly through recurrent neural networks (RNNs) such as long short-term memory (LSTM) architectures, has transformed the analysis of trends in vulnerability discovery and exploitation. These models address the limitations of static approaches by effectively capturing temporal dependencies and sequential relationships within data. Ensemble learning methods, including gradient boosting and bagging, have also demonstrated strong potential in improving predictive performance by integrating the outputs of multiple models to minimize variance and bias [8] [9] [10]. Despite recent progress, key challenges remain in predictive cybersecurity. Data imbalance often limits model generalization, while evolving threat landscapes demand frequent retraining to preserve accuracy. The opacity of deep learning models also complicates interpretability, reducing analyst confidence. Effective evaluation must therefore incorporate comprehensive metrics such as precision, recall, F1-score and AUC to ensure real-world applicability and reliability.

Ensemble learning has emerged as a central focus in contemporary cybersecurity research due to its capacity to enhance predictive accuracy and model robustness through the integration of multiple machine learning algorithms. Unlike single-model approaches, ensemble techniques such as bagging, boosting and stacking harness the complementary strengths of different base learners to mitigate individual weaknesses, reduce variance and improve generalization. Recent advancements extend these principles into the deep learning domain, where ensembles of specialized neural networks are combined either through averaging mechanisms or trainable meta-models to capture complex nonlinear patterns and temporal dependencies in vulnerability data. Empirical studies indicate that such ensemble deep learning frameworks outperform traditional single-model approaches in predicting exploitable network vulnerabilities, offering superior adaptability to data noise, class imbalance and diverse feature importance [11] [12]. Furthermore, emerging architectures like Divergent Ensemble Networks demonstrate an optimal balance between computational efficiency and predictive diversity by employing shared feature representations alongside independent branches to improve uncertainty estimation. This integration of ensemble methodologies with deep neural architectures signals a promising direction for developing reliable, interpretable and scalable systems capable of forecasting vulnerabilities in dynamic cybersecurity environments.

III. Methodology

Developing predictive machine learning models for forecasting network vulnerabilities involves a sequential process encompassing data collection, preprocessing, and feature engineering. Data are gathered from credible sources such as the National Vulnerability Database (NVD), Exploit DB, and various security advisories, capturing both static vulnerability attributes and dynamic exploitation patterns. Preprocessing ensures data quality through imputation, normalization, and categorical encoding, while imbalanced datasets are addressed using techniques like SMOTE and weighted loss functions. Feature engineering focuses on extracting key predictors, including CVSS metrics such as attack vector, complexity, privileges and impact on confidentiality, integrity, and availability, alongside features derived from textual descriptions and network context through natural language processing and temporal analysis. Figure 1 shows predictive vulnerability modeling process.

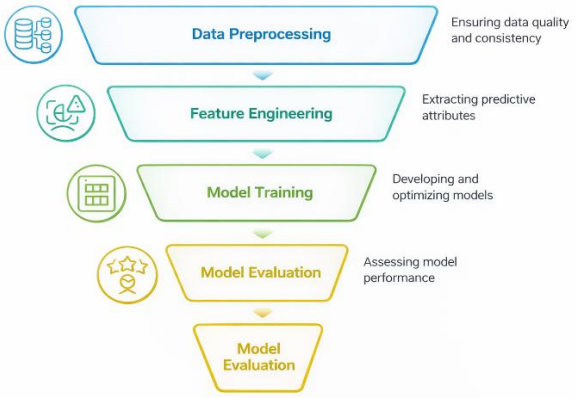

Fig 1: Predictive Vulnerability Modeling Process

A diverse set of supervised machine learning algorithms is employed to address the challenge of predicting vulnerability exploitation. These include random forests, gradient boosting machines, support vector machines and deep neural networks. To optimize performance and prevent overfitting, systematic techniques such as cross-validation and hyperparameter tuning are applied. Considering the temporal nature of cyber threats, sequential models like long short-term memory (LSTM) networks are utilized to capture temporal dependencies and evolving patterns within vulnerability data. Model interpretability is achieved through methods such as SHAP (SHapley Additive exPlanations), which identify the most influential features contributing to each prediction, thereby enhancing transparency and facilitating actionable decision-making. Model performance is evaluated using a comprehensive set of metrics, including precision, recall, F1-score and the area under the receiver operating characteristic (ROC) curve (AUC). The overarching goal is to maintain an effective balance between false positives and false negatives, ensuring the practical applicability of these models for vulnerability prioritization and the formulation of proactive

mitigation strategies. Figure 2 shows machine learning model optimization.

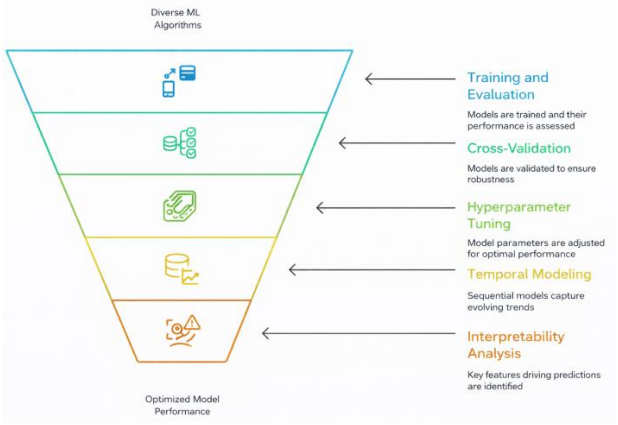

Fig 2: Machine Learning Model Optimization Funnel

IV. Experimental Results and Analysis

The experimental phase focuses on evaluating the proposed predictive models using benchmark datasets derived from publicly available vulnerability and exploit repositories. Multiple machine learning classifiers were implemented and tested to assess their effectiveness in predicting which vulnerabilities are most likely to be exploited within a given timeframe. The findings indicate that ensemble-based approaches, such as random forests and gradient boosting, deliver superior overall performance, particularly with respect to precision and recall. Figure 3 shows performance comparison of predictive models for vulnerability forecasting.

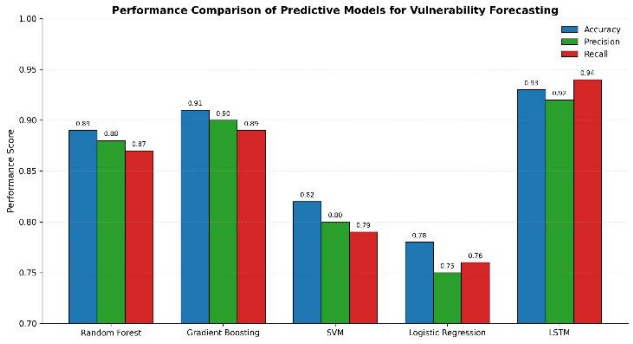

Fig 3: Performance Comparison of Predictive Models for Vulnerability Forecasting

Deep learning models, especially long short-term memory (LSTM) networks, exhibit a strong capacity to capture temporal dependencies and evolving trends in exploitation activities, thereby improving short and medium-term forecasting accuracy. A grouped bar chart illustrates the comparative performance of the tested models across key metrics, including accuracy, precision and recall. Ensemble methods such as Random Forest and Gradient Boosting consistently achieve competitive results across all indicators, while the LSTM model demonstrates an exceptional ability to model temporal dynamics, attaining the highest overall scores. The models evaluated in this study include Random Forest, Gradient Boosting, Support Vector Machine (SVM), Logistic Regression and LSTM networks.

Addressing data imbalance through techniques such as the Synthetic Minority Over-sampling Technique (SMOTE) and class-weighted loss functions significantly enhanced the detection of exploitable vulnerabilities without inflating false alarm rates. Analysis of feature importance revealed that attributes such as attack complexity, user interaction requirements, and historical exploitation patterns served as strong predictors of exploitability. The models also generated valuable insights into emerging exploit trends and high-risk vulnerabilities, enabling more adaptive threat assessments and informed resource prioritization. Comparative evaluations highlighted inherent trade-offs between model complexity, interpretability and computational cost, demonstrating the operational effectiveness of tree-based algorithms in various deployment scenarios. The experimental findings reinforced the capacity of machine learning classifiers to identify exploitable network vulnerabilities, revealing notable variations in predictive performance and applicability. Ensemble methods, particularly random forests and gradient boosting, emerged as top-performing models, delivering high accuracy while maintaining balanced precision and recall. These models efficiently handled heterogeneous feature sets and mitigated the effects of data noise prevalent in vulnerability repositories. Furthermore, their strong area under the receiver operating characteristic (ROC) curve (AUC) confirmed their robustness in discriminating between exploitable and non-exploitable vulnerabilities across diverse datasets and temporal contexts.

Deep learning models, particularly long short-term memory (LSTM) networks, have demonstrated exceptional capability in capturing temporal dependencies within vulnerability exploitation data. These architectures outperform traditional classifiers by accurately forecasting exploitation trends over short and medium-term intervals, effectively modeling the sequential evolution of attack patterns and vulnerability emergence over time. Their inherent ability to retain contextual information and temporal relationships allows LSTM networks to adapt effectively to the rapidly changing threat environment. In addition to predictive accuracy, these models offer valuable interpretive insights by highlighting vulnerabilities at greater risk of exploitation, thereby allowing cybersecurity teams to prioritize mitigation efforts more strategically. By integrating sequence modeling with feature importance analysis, it becomes possible to identify specific attributes and behavioral patterns that correlate with higher exploitability, transforming predictions into actionable intelligence. A grouped bar chart illustrating the forecasting performance of LSTM networks relative to traditional models across multiple time horizons; short-term (1 week), mid-term (1 month), and long-term (3 months) shows that LSTMs consistently achieve superior results. This performance advantage underscores the strength of temporal modeling in uncovering hidden dependencies and emerging trends in vulnerability exploitation dynamics. Figure 4 shows temporal forecasting performance comparison of LSTM and traditional models.

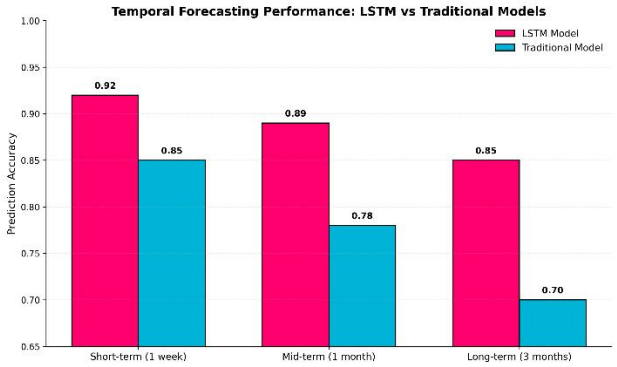

Fig 4: Temporal Forecasting Performance: LSTM vs. Traditional Models

Addressing class imbalance proved essential for enhancing both the sensitivity and specificity of the predictive models. Techniques such as the Synthetic Minority Over-sampling Technique (SMOTE) and class weighting played a significant role in improving the detection of exploited vulnerabilities, which, although relatively rare, represent the most critical security threats. These resampling strategies balanced the training data distribution, enabling the models to better capture patterns associated with exploitable vulnerabilities without overfitting to the majority class. Consequently, the models achieved higher detection rates for high-risk vulnerabilities while reducing the likelihood of overlooking critical exploits. A minor increase in false positive rates was observed, underscoring the importance of parameter optimization and model calibration. Overall, the results highlight that effectively mitigating class imbalance is fundamental to developing robust and trustworthy predictive frameworks for real-world cybersecurity applications.

Feature importance analysis played a crucial role in evaluating the predictive capabilities of the models. Variables such as attack complexity, required privileges, and historical exploitation data emerged as the most influential factors in determining exploitability. These findings underscore the critical role of attack vectors and prior exploit occurrences in vulnerability forecasting, contributing to improved interpretability and informed security decision-making. By identifying these key attributes, analysts can direct their efforts toward vulnerabilities that exhibit a higher likelihood of exploitation, thereby streamlining patch management and mitigation planning. Furthermore, the models' ability to reveal influential predictive features enhances both transparency and trust, which are vital for successful operational integration. The experimental outcomes demonstrate that combining high predictive accuracy with interpretability significantly strengthens the practical applicability and reliability of vulnerability forecasting systems.

The comparative assessment of different modeling approaches revealed several critical trade-offs in performance and practicality. Deep learning models demonstrated exceptional predictive accuracy but required substantial computational resources and extended training durations, posing challenges for deployment in real-time or resource-constrained environments. In contrast, tree-based ensemble methods provided a more balanced alternative, coupling robust predictive capability with lower complexity and greater interpretability. The findings also indicated that increasing model complexity amplifies the difficulties of operational implementation and system integration. Consequently, model selection should carefully consider factors such as computational infrastructure, latency constraints and explainability requirements. These insights emphasize the importance of aligning modeling strategies with an organization's technical capacity and security objectives, ensuring that predictive systems remain both effective and trustworthy within operational contexts.

The experimental evaluations demonstrate that predictive machine learning models are highly effective in identifying exploitable vulnerabilities, offering substantial potential to strengthen cybersecurity operations. The observed improvements in accuracy, robustness and interpretability underscore a transformative shift from reactive vulnerability management focused on post-exploitation mitigation to proactive strategies that anticipate and address threats in advance. Despite these advances, several challenges remain, including the need to maintain high data quality, mitigate class imbalance and ensure continuous model adaptation to the rapidly evolving threat environment. For organizations adopting these systems, successful implementation will depend on seamless integration with existing security frameworks and the automation of response processes to achieve operational efficiency. Future research should focus on enhancing model scalability, improving resilience against adversarial tactics, and enabling real-time deployment within dynamic network contexts. Overall, the findings reaffirm that predictive machine learning has the potential to transform traditional vulnerability management into a proactive, intelligence-driven discipline that reduces organizational risk and strengthens cybersecurity resilience.

V. Discussion

The findings highlight both the potential and the challenges of applying predictive machine learning to vulnerability forecasting in operational settings. Ensemble methods enhance accuracy and robustness by combining diverse base learners, reducing overfitting and improving stability in complex cybersecurity data. However, severe class imbalance remains a critical issue, requiring adaptive sampling or cost-sensitive strategies to maintain sensitivity to rare but high-impact vulnerabilities. Ensuring interpretability through techniques such as SHAP further supports analyst trust, while balancing detection performance and false positive rates remains essential for effective deployment. The evolving nature of cybersecurity requires predictive models to be continuously retrained with current vulnerability data and emerging threat intelligence to remain effective. Integration with real-time security frameworks enhances context-aware risk assessment, while model interpretability through techniques such as feature importance analysis and SHAP strengthens transparency and practitioner trust. As adversarial attacks grow more sophisticated, incorporating adversarial training, anomaly

detection, and feedback mechanisms improves system robustness. Attention to data privacy and secure data management further ensures ethical and reliable deployment within dynamic security environments.

Integrating predictive vulnerability models into existing cybersecurity frameworks presents a range of both technical and organizational challenges. From a technical standpoint, these systems must seamlessly interface with security information and event management (SIEM) tools, vulnerability management solutions, and incident response workflows to generate timely and actionable intelligence. Effective integration ensures that predictive outputs translate into operational value by enhancing situational awareness and guiding proactive defense measures. On the organizational side, the establishment of well-defined roles, responsibilities, and decision-making protocols is essential to govern automated and semi-automated processes for vulnerability prioritization. Moreover, the ethical and privacy implications of handling sensitive network and threat intelligence data require stringent compliance measures and secure management practices. Maintaining transparency and adherence to data protection standards is critical for safeguarding stakeholder trust and ensuring responsible deployment within operational cybersecurity environments. Figure 5 shows key factors in successful integration of predictive vulnerability models.

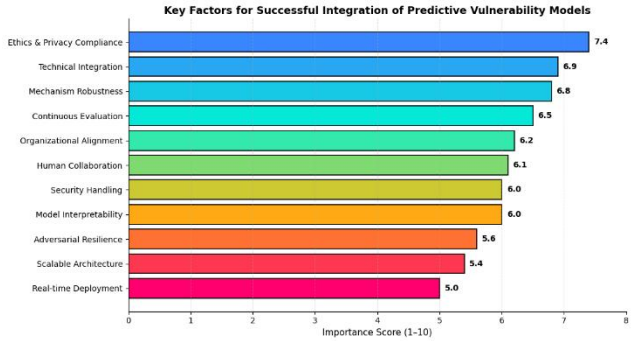

Fig 5: Key Factors in Successful Integration of Predictive Vulnerability Models

In conclusion, the successful implementation of predictive vulnerability models requires a holistic approach that combines technological optimization, operational integration, continuous evaluation and active human collaboration. These interconnected components collectively determine the effectiveness and sustainability of forecasting systems within dynamic cybersecurity environments. Predictive machine learning models, particularly those leveraging ensemble techniques, hold significant potential to transform proactive cybersecurity by delivering timely and data-driven insights into emerging vulnerabilities. However, their long-term success depends on a multidisciplinary framework that unites technical efficiency, interpretability, continuous adaptation, security reinforcement, and ethical governance. Future research should focus on developing scalable architectures, enhancing robustness against adversarial manipulation, and achieving real-time deployment in complex operational contexts. A horizontal bar chart summarizing the key enablers of successful model integration underscores the relative importance of factors such as technical interoperability, organizational cohesion, iterative assessment, human collaboration, privacy compliance, model robustness, explainability, system hardening, scalability, adversarial resistance and real-time adaptability.

VI. Conclusion

This paper explores the pivotal role of predictive machine learning models in forecasting network vulnerabilities that are susceptible to exploitation. It emphasizes the transformative potential of shifting cybersecurity paradigms from reactive defense to proactive threat anticipation. By leveraging advanced algorithms such as ensemble learning techniques and deep learning architectures, these models enable accurate identification of high-risk vulnerabilities, providing actionable insights for strategic defense planning. The proposed framework spanning comprehensive data acquisition, advanced feature engineering, and model interpretability establishes a foundation for the development of resilient forecasting systems. Experimental findings validate the effectiveness of the models while highlighting key challenges such as class imbalance, temporal dynamics and data heterogeneity. Although obstacles related to data quality, model transparency and operational integration persist, predictive machine learning represents a substantial advancement in strengthening cybersecurity resilience. Future research should focus on incorporating real-time intelligence, enhancing robustness against adversarial manipulation, and facilitating automated response mechanisms guided by predictive alerts. Ultimately, integrating predictive modeling into vulnerability management processes presents a promising approach to minimizing attack surfaces, prioritizing remediation efforts and fortifying digital infrastructures against an increasingly sophisticated and dynamic threat landscape.

VII. Future Work

Future research on predictive machine learning for vulnerability forecasting should prioritize real-time adaptability and closer integration with dynamic threat intelligence to address the growing volume and complexity of cyberattacks fueled by AI advancements. Developing hybrid models that combine traditional learning approaches with generative AI can improve simulation of multi-stage exploits and enhance proactive defense strategies. Strengthening ties with zero trust frameworks and continuous monitoring systems will enable automated, real-time threat mitigation. Moreover, ensuring robustness against adversarial manipulation and maintaining interpretability will be essential for sustaining analyst confidence. Advancements in privacy-preserving and scalable cloud-native frameworks will further support real-time processing of large cybersecurity datasets, promoting widespread adoption and operational impact in safeguarding digital infrastructure.

References

[1] I. H. Sarker, "Machine learning for intelligent data analysis and automation in cybersecurity: current and

future prospects," *Annals of Data Science,* vol. 10, no. 6, pp. 1473-1498, December 2023.

[2] N. Sun, J. Zhang, P. Rimba, S. Gao, L. Y. Zhang and Y. Xiang, "Data-Driven Cybersecurity Incident Prediction: A Survey," *IEEE Communications Surveys & Tutorials,* vol. 21, no. 2, pp. 1744-1772, 2019.

[3] G. Jabeen, S. Rahim, W. Afzal, D. Khan, A. A. Khan, Z. Hussain and T. Bibi, "Machine learning techniques for software vulnerability prediction: a comparative study," *Applied Intelligence,* vol. 52, no. 15, pp. 17614-17635, December 2022.

[4] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan and R. Jain, "Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things," *IEEE Internet of Things Journal,* vol. 6, no. 4, pp. 6822-6834, 2019.

[5] F. Alenezi and C. P. Tsokos, "Machine Learning Approach to Predict Computer Operating Systems Vulnerabilities," in *3rd International Conference on Computer Applications & Information Security (ICCAIS)*, Riyadh, Saudi Arabia, 2020.

[6] Y. Ma, T. Xie, J. Li and R. Maciejewski, "Explaining Vulnerabilities to Adversarial Machine Learning through Visual Analytics," *IEEE Transactions on Visualization and Computer Graphics,* vol. 26, no. 1, pp. 1075-1085, 2020.

[7] J. Malik, R. Muthalagu and P. M. Pawar, "A Systematic Review of Adversarial Machine Learning Attacks, Defensive Controls, and Technologies," *IEEE Access,* vol. 12, no. 1, pp. 99382-99421, 2024.

[8] H. HaddadPajouh, A. Dehghantanha, R. Khayami and K.-K. R. Choo, "A deep Recurrent Neural Network based approach for Internet of Things malware threat hunting," *Future Generation Computer Systems,* vol. 85, no. 1, pp. 88-96, 2018.

[9] I. H. Sarker, "Deep Cybersecurity: A Comprehensive Overview from Neural Network and Deep Learning Perspective," *SN Computer Science,* vol. 2, no. 3, p. 154, March 2021.

[10] P. Zeng, G. Lin, L. Pan, Y. Tai and J. Zhang, "Software Vulnerability Analysis and Discovery Using Deep Learning Techniques: A Survey," *IEEE Access,* vol. 8, pp. 197158-197172, 2020.

[11] S. Abimannan, E.-S. M. El-Alfy, Y.-S. Chang, S. Hussain, S. Shukla and D. Satheesh, "Ensemble Multifeatured Deep Learning Models and Applications: A Survey," *IEEE Access,* vol. 11, pp. 107194-107217, 2023.

[12] M. Sakib, S. Mustajab and M. Alam, "Ensemble deep learning techniques for time series analysis: a comprehensive review, applications, open issues, challenges, and future directions," *Cluster Computing,* vol. 28, no. 1, p. 73, 2024.