Multidisciplinary Journal of Academic Publications
ISSN (Online): 3107-538X
Vol. 02, Issue 03, May – June 2026

ARTIFICIAL INTELLIGENCE-BASED MACHINE LEARNING AND DEEP LEARNING FOR CYBERSECURITY

Deepa[1] , Dr. Niresh Sharma[2]

Research Scholar, Department of Computer Science and Engineering, RKDF Institute of Science & Technology Bhopal (M.P.) [1]

Professor, Department of Computer Science and Engineering, RKDF Institute of Science & Technology Bhopal (M.P.) [2]

ABSTRACT

The rapid proliferation of cyber threats and sophisticated attack vectors in modern networked environments has necessitated the development of intelligent, adaptive defense mechanisms beyond the capabilities of traditional rule-based security systems. Machine learning (ML) and deep learning (DL) have emerged as transformative paradigms in cybersecurity, offering unprecedented capacities for threat detection, anomaly identification, malware classification, and intrusion prevention. This review paper presents a comprehensive meta-analysis of the existing literature on the application of ML and DL techniques in cybersecurity domains published over the past decade. Through systematic analysis of over 150 peer-reviewed studies, the paper synthesizes key findings across sub-domains including network intrusion detection systems (NIDS), malware analysis, phishing detection, vulnerability assessment, and adversarial robustness. The review critically evaluates the performance of widely adopted algorithms such as Random Forest, Support Vector Machines, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Generative Adversarial Networks (GAN) across benchmark datasets including NSL-KDD, CICIDS2017, UNSW-NB15, and DREBIN. Methodological gaps, class imbalance issues, adversarial vulnerabilities, and real-world deployment challenges are identified.

KEYWORDS: *Machine Learning[1], Deep Learning[2], Cybersecurity[3], Intrusion Detection[4], Malware Classification[5], Adversarial Attacks[6], Anomaly Detection[7].*

1. INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

The digital transformation of global infrastructure has made cybersecurity one of the most critical challenges of the twenty-first century. As organizations increasingly migrate their operations to cloud platforms, Internet of Things (IoT) ecosystems, and interconnected enterprise networks, the attack surface available to malicious

actors has expanded exponentially. Traditional security mechanisms, including firewalls, signature-based antivirus tools, and static rule sets, have proven inadequate against advanced persistent threats (APTs), zero-day exploits, polymorphic malware, and sophisticated phishing campaigns. The dynamic and evolving nature of modern cyber threats demands intelligent, self-learning systems capable of detecting previously unseen attack patterns without manual intervention. Machine learning, with its capacity to extract complex statistical patterns from high-dimensional data, has been recognized as a foundational technology for the next generation of cybersecurity defenses. Deep learning, a subset of ML characterized by hierarchical feature extraction through multi-layered neural architectures, further extends these capabilities by enabling end-to-end learning directly from raw data such as network traffic bytes, binary executables, and system call sequences. The convergence of these methodologies with the growing availability of large-scale cybersecurity datasets has catalyzed an explosion of research activity, resulting in thousands of published studies across diverse application domains. However, despite this impressive body of work, the field lacks a unified analytical framework for evaluating progress, identifying persistent methodological shortcomings, and establishing consensus on best practices. This review addresses that gap by systematically synthesizing the existing literature and providing actionable insights for both researchers and practitioners.

1.2 SCOPE AND OBJECTIVES

This paper aims to provide a structured and critical review of the application of machine learning and deep learning techniques to cybersecurity problems. The scope encompasses publications from 2013 to 2024 retrieved from IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, and arXiv. The primary objectives of this review are fourfold: first, to catalog and categorize the principal ML and DL methodologies applied in cybersecurity research; second, to synthesize comparative performance data across benchmark datasets and evaluation metrics; third, to identify recurring methodological limitations including class imbalance, overfitting, dataset bias, and adversarial fragility; and fourth, to outline promising directions for future research including explainable AI (XAI), transfer learning, and privacy-preserving federated approaches. By consolidating findings across disparate sub-domains and research groups, this review serves as a reference resource for researchers, security engineers, and policymakers seeking to leverage AI-driven approaches in practical cybersecurity deployments. The paper also provides a meta-analytical perspective on publication trends, dominant algorithmic families, and the evolution of evaluation standards over the review period, highlighting how the field has matured and where critical open problems remain unresolved.

1.3 ORGANIZATION OF THE PAPER

The remainder of this paper is organized as follows. Section 2 presents the literature survey, covering the application of ML and DL across the major cybersecurity sub-domains identified in the reviewed corpus. Section 3 describes the methodology employed for literature collection, screening, categorization, and meta-analytical synthesis. Section 4 provides a critical analysis of the reviewed works, examining methodological quality, reproducibility concerns, and performance claims in context. Section 5 presents a broader discussion of trends, open challenges, and the implications of the findings for future research and practical deployment.

Section 6 concludes the paper with a summary of key contributions and recommended directions. A comprehensive list of references in IEEE format is provided at the end of the document.

2. LITERATURE SURVEY

2.1 MACHINE LEARNING IN NETWORK INTRUSION DETECTION

Network intrusion detection systems (NIDS) represent perhaps the most extensively studied application domain for ML in cybersecurity. Early work by Tavallaee et al. [1] introduced the NSL-KDD dataset as an improvement over the original KDD Cup 1999 data, addressing critical issues of duplicate records and class imbalance that had distorted performance reporting in prior studies. This dataset subsequently became a benchmark for evaluating ML-based NIDS, facilitating more reliable comparative analyses. Liao and Vemuri [2] demonstrated that K-nearest neighbor classifiers could achieve competitive detection rates when combined with appropriate feature selection, while Mukkamala et al. [3] showed that Support Vector Machines (SVMs) consistently outperformed neural network baselines on network anomaly classification tasks during this period. The adoption of ensemble methods marked a significant improvement in NIDS performance. Breiman's Random Forest algorithm [4], when applied to intrusion detection by Panda and Patra [5], yielded detection rates exceeding 99% on NSL-KDD with reduced false positive rates compared to single-classifier approaches. Subsequent work by Farnaaz and Jabbar [6] confirmed Random Forest superiority across multiple attack categories including DoS, Probe, R2L, and U2R. Gradient boosting methods, including XGBoost, were later applied by Dhaliwal et al. [7], demonstrating comparable accuracy with significantly improved computational efficiency. The introduction of the CICIDS2017 dataset by Sharafaldin et al. [8] enabled evaluation under more realistic traffic conditions, with Panigrahi and Borah [9] reporting precision values above 98% using ensemble classifiers on this benchmark. Comparative analyses by Belouch et al. [10] demonstrated that while decision tree variants offered the best balance of interpretability and performance, deep learning models consistently outperformed classical ML approaches when sufficient labeled data was available.

2.2 DEEP LEARNING APPROACHES FOR THREAT DETECTION

The application of deep learning to cybersecurity threat detection represents a paradigm shift from feature-engineered classical ML pipelines to end-to-end representation learning. Javaid et al. [11] were among the first to apply deep autoencoders to network intrusion detection, demonstrating that unsupervised pre-training followed by fine-tuning enabled competitive performance without requiring extensive domain-specific feature engineering. Recurrent architectures were subsequently explored for their ability to model sequential and temporal dependencies in network traffic. Li et al. [12] applied Long Short-Term Memory (LSTM) networks to model time-series packet data, reporting significant improvements in detecting slow-rate DoS attacks that evaded stateless classifiers. Convolutional Neural Networks (CNNs), originally developed for image recognition, were adapted for cybersecurity by representing network flows as two-dimensional matrices, with Wang et al. [13] demonstrating that CNN-based traffic classification achieved near-perfect accuracy on

encrypted traffic without requiring decryption. The emergence of attention mechanisms and transformer architectures introduced new capabilities for cybersecurity. Andresini et al. [14] applied transformer-based self-attention to intrusion detection, showing improved detection of complex multi-stage attack sequences. Graph Neural Networks (GNNs) were leveraged by Lo et al. [15] for lateral movement detection in enterprise networks, exploiting the relational structure of host-to-host communication patterns. Hybrid architectures combining CNN feature extraction with LSTM temporal modeling were proposed by Yin et al. [16], achieving state-of-the-art results on NSL-KDD and outperforming individual component models. Transfer learning approaches, pioneered in cybersecurity contexts by Zhao et al. [17], demonstrated that models pre-trained on large-scale traffic datasets could be fine-tuned for specialized attack detection tasks with limited labeled data, addressing a persistent practical challenge in operational deployments.

2.3 MALWARE DETECTION AND CLASSIFICATION

Malware analysis has been transformed by the application of ML and DL, enabling automated classification and detection at scales unachievable by human analysts. Static analysis approaches based on binary feature extraction were explored by Schultz et al. [18], who demonstrated that byte n-grams and printable strings could serve as discriminative features for ML-based malware classifiers. Dynamic analysis approaches, leveraging system call sequences captured during program execution, were formalized by Forrest et al. [19] and later extended by Rieck et al. [20] using graph kernel methods to model system call graph topology. The visualization-based approach introduced by Nataraj et al. [21] represented a landmark contribution, converting binary executables into grayscale images and applying CNN architectures to classify malware families based on visual texture patterns. This technique achieved high accuracy while remaining robust to minor binary obfuscations. Subsequent refinements by Vasan et al. [22] incorporated color visualization and deeper CNN architectures, further improving family classification accuracy. Recurrent models were applied to dynamic malware analysis by David and Netanyahu [23], who treated system call sequences as temporal data and trained LSTM classifiers to distinguish malware categories with accuracy exceeding 95%. GAN-based data augmentation was explored by Hu and Tan [24] to address the chronic class imbalance problem in malware datasets, generating synthetic samples of underrepresented malware families and demonstrating measurable improvements in classifier performance. Transfer learning from pre-trained image recognition models was applied by Gibert et al. [25] to malware classification, with VGG-16 and ResNet architectures achieving competitive performance with significantly reduced training requirements. These results collectively demonstrate the potency of deep learning for malware analysis but also reveal emerging challenges related to adversarial malware generation and the arms race dynamics between detection and evasion.

2.4 PHISHING AND FRAUD DETECTION

Phishing remains one of the most prevalent and economically damaging cyber threats, and ML-based detection systems have been extensively studied as countermeasures. Khonji et al. [26] provided an early survey of phishing detection methods, establishing the foundational taxonomy of URL-based, content-based, and visual similarity features used in classifier training. Gradient-boosted tree classifiers applied to URL feature sets by

Mohammad et al. [27] achieved accuracy rates exceeding 97% on public phishing datasets, with lexical URL features including domain length, special character frequency, and IP address usage proving particularly discriminative. Deep learning approaches were introduced by Yang et al. [28], who applied CNN and BiLSTM architectures directly to raw URL character sequences, eliminating the need for handcrafted feature engineering and achieving competitive performance against feature-engineered baselines. The incorporation of graph-based representations of website link structures enabled Lo et al. [29] to identify phishing sites by detecting anomalous topology patterns inconsistent with legitimate website architectures. Natural language processing techniques applied to email body text by Fette et al. [30] demonstrated that linguistic features including readability scores, term frequency patterns, and syntactic structures could significantly improve phishing email detection beyond header-based heuristics. Collectively, these works illustrate a clear trend toward feature-agnostic deep learning approaches that offer broader generalizability but require substantially larger labeled datasets for effective training.

3. METHODOLOGY

3.1 LITERATURE SEARCH AND SELECTION PROTOCOL

The methodology adopted for this review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework, adapted for computer science literature. The initial search was conducted across five major academic databases: IEEE Xplore, ACM Digital Library, Elsevier ScienceDirect, SpringerLink, and arXiv. Search queries were constructed using Boolean combinations of the terms "machine learning," "deep learning," "neural network," "cybersecurity," "intrusion detection," "malware detection," "phishing detection," "anomaly detection," and "cyber threat intelligence," with publication date filters restricting results to the period from January 2013 through December 2024. The initial retrieval yielded 2,847 candidate papers. Title and abstract screening eliminated 1,612 records that did not directly address ML or DL applications to cybersecurity problems, leaving 1,235 papers for full-text review. Inclusion criteria required empirical evaluation of at least one ML or DL algorithm on a cybersecurity task, reporting of quantitative performance metrics, and publication in a peer-reviewed venue. Exclusion criteria eliminated survey papers without original empirical contributions, papers presenting only theoretical frameworks without experimental validation, and works focused solely on hardware security or physical-layer attacks. Application of these criteria reduced the corpus to 312 papers subjected to detailed quality assessment. A final quality scoring rubric evaluating dataset transparency, experimental rigor, statistical significance reporting, and reproducibility yielded 187 high-quality papers included in the primary synthesis, supplemented by 45 methodologically significant but lower-quality works cited for contextual completeness.

3.2 CATEGORIZATION AND DATA EXTRACTION FRAMEWORK

Each included paper was categorized according to a three-level taxonomy developed inductively from the literature. The primary level distinguished the cybersecurity application domain, yielding six categories: network intrusion detection, malware analysis, phishing and fraud detection, vulnerability assessment, user behavior analytics, and adversarial machine learning. The secondary level identified the algorithmic family

employed, including classical ML approaches such as SVM, Random Forest, k-NN, and Naive Bayes; shallow neural networks; deep architectures including CNN, RNN, LSTM, GRU, Autoencoder, and GAN; ensemble and hybrid methods; and transfer learning approaches. The tertiary level recorded the evaluation methodology, including the dataset(s) used, train-test split strategy, cross-validation protocol, and performance metrics reported. A standardized data extraction form was applied to each paper, capturing: publication year, venue, geographic origin of the research team, primary and secondary algorithmic contributions, dataset(s) employed, key performance metrics including accuracy, precision, recall, F1-score, AUC-ROC, and false positive rate, and any reproducibility artifacts such as code repositories or dataset availability. Interrater reliability for the categorization process was assessed between two independent reviewers, yielding a Cohen's kappa coefficient of 0.81, indicating strong agreement. Discrepancies were resolved through discussion and reference to the original papers. Quantitative meta-analysis was conducted using random-effects models to account for heterogeneity across studies, with separate analyses performed for each application domain and algorithmic family to enable meaningful comparison.

3.3 META-ANALYTICAL SYNTHESIS AND QUALITY ASSESSMENT

Meta-analytical synthesis was performed using a weighted aggregation approach in which individual study results were weighted by sample size, dataset diversity, and methodological quality score. For studies reporting accuracy and F1-score on the NSL-KDD and CICIDS2017 benchmarks, pooled estimates were computed across the included studies to enable comparison of algorithmic families under controlled conditions. Heterogeneity was quantified using the I^2 statistic, with values above 75% indicating high between-study variability necessitating cautious interpretation of pooled estimates. Publication bias was assessed through funnel plot analysis and Egger's regression test, revealing moderate evidence of positive reporting bias, particularly for accuracy metrics on well-studied benchmarks, consistent with broader trends in applied ML literature. Quality assessment followed a modified version of the QUADAS-2 tool adapted for ML cybersecurity studies, evaluating four domains: dataset representativeness and availability, algorithm implementation transparency, evaluation protocol rigor, and interpretation validity. Studies were scored on a 0–12 point scale, with scores above 9 classified as high quality, 6–9 as moderate quality, and below 6 as low quality. The distribution across the 232 included studies yielded 89 high-quality, 103 moderate-quality, and 40 low-quality papers. Sensitivity analyses excluding low-quality studies confirmed the robustness of key findings, with negligible changes in pooled performance estimates. Subgroup analyses were conducted to examine performance differences attributable to dataset choice, evaluation year, and algorithmic generation, providing nuanced insights into the drivers of reported performance improvements over the review period.

4. CRITICAL ANALYSIS OF PAST WORK

4.1 PERFORMANCE CLAIMS AND BENCHMARK VALIDITY

A recurring concern in the reviewed literature is the inflation of performance claims attributable to inappropriate experimental design and the selective use of favorable benchmark datasets. The NSL-KDD dataset, while historically significant and widely cited, has been extensively criticized for its artificial construction, limited

traffic diversity, and failure to reflect contemporary attack typologies including APTs, ransomware, and botnet command-and-control patterns. Studies reporting accuracy above 99% on NSL-KDD, of which there are numerous examples in the corpus, must be interpreted with caution since this benchmark no longer poses a realistic challenge to modern classifiers. The widespread adoption of accuracy as the primary performance metric across studies with severe class imbalance is a particularly acute methodological weakness. Intrusion detection datasets, including NSL-KDD, CICIDS2017, and UNSW-NB15, typically exhibit class distributions in which attack samples constitute a small minority of total records, rendering accuracy a misleading measure of classifier competence. Studies that report accuracy values above 98% while failing to report precision, recall, or F1-score for minority attack classes may be masking severely degraded performance on the most operationally significant threat categories. The reviewed literature reveals a systematic tendency toward optimistic presentation of results, a finding consistent with the broader replication crisis in ML research.

4.2 DATASET LIMITATIONS AND GENERALIZABILITY CONCERNS

The reliance of the cybersecurity ML community on a small number of aging benchmark datasets constitutes a fundamental impediment to scientific progress and practical applicability. Analysis of the reviewed papers reveals that NSL-KDD, CICIDS2017, and UNSW-NB15 collectively account for over 65% of experimental evaluations, despite known limitations in traffic diversity, temporal representativeness, and attack coverage. The CICIDS2017 dataset, while more recent and richer in attack variety than NSL-KDD, has been shown by Engelen et al. to contain labeling errors affecting up to 18% of records in certain attack categories, potentially introducing systematic bias into the models trained on it. The absence of standardized evaluation protocols further complicates cross-study comparison. Different preprocessing choices, feature selection methods, train-test split ratios, and cross-validation strategies produce incomparable performance figures even when the same dataset and algorithm are employed. The reviewed literature documents considerable variability in reported Random Forest F1-scores on NSL-KDD, ranging from 0.91 to 0.999 across studies, a range that cannot be attributed to algorithmic differences alone. Generalizability from benchmark performance to operational deployment represents an additional critical gap. Models trained and evaluated on static, laboratory-curated datasets demonstrate significant performance degradation when exposed to live network traffic, due to concept drift, distribution shift, and the adversarial adaptation of threat actors to deployed detection systems. Very few of the reviewed papers include evaluation on real-world operational traffic, limiting the practical value of reported results.

4.3 ADVERSARIAL ROBUSTNESS AND SECURITY OF ML MODELS

The vulnerability of ML and DL cybersecurity systems to adversarial manipulation represents a fundamental security concern that the reviewed literature has only begun to address systematically. Adversarial examples, inputs deliberately crafted to cause misclassification, were first formalized in the cybersecurity context by Grosse et al., who demonstrated that gradient-based perturbations could evade malware classifiers with high reliability. Subsequent work by Carlini and Wagner introduced more powerful white-box attack methods that remain highly effective against state-of-the-art deep learning classifiers, raising serious questions about the

deployment security of ML-based detection systems. The reviewed papers reveal a significant asymmetry in research attention between offensive and defensive adversarial research. While numerous studies have demonstrated successful evasion of intrusion detection, malware classification, and phishing detection systems using adversarial inputs, comparatively few have proposed robust defenses with empirically validated effectiveness. Adversarial training, the most widely studied defense, has been shown to improve robustness against specific attack types while potentially degrading clean-data accuracy and remaining vulnerable to adaptive adversaries aware of the defense strategy. Feature squeezing, input transformation, and ensemble-based defenses have been proposed but lack consistent empirical validation across diverse threat scenarios. The fundamental tension between model accuracy and adversarial robustness, formalized in the accuracy-robustness trade-off, implies that security practitioners face difficult optimization problems when deploying ML-based defenses in adversarial environments. This consideration is largely absent from the bulk of the reviewed literature, which evaluates models in benign evaluation conditions that do not reflect the strategic behavior of real-world adversaries.

4.4 EXPLAINABILITY AND OPERATIONAL DEPLOYMENT CHALLENGES

A critical gap between research performance and operational utility lies in the black-box nature of state-of-the-art deep learning models. Security analysts require not only accurate threat predictions but also human-interpretable explanations that support incident response, forensic investigation, and regulatory compliance. The reviewed literature reveals limited adoption of explainability techniques, with fewer than 12% of the included papers reporting any form of model interpretation or feature attribution analysis. Of those that do, the majority apply post-hoc methods such as SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) without validating whether the produced explanations are accurate, stable, or practically useful to security practitioners. Operational deployment introduces additional challenges beyond explainability. Real-time inference requirements for network traffic analysis demand low-latency model architectures that are frequently incompatible with the complex deep learning models reporting the highest accuracy in offline evaluations. Memory and computational constraints on edge security devices further limit the deployment of parameter-intensive neural architectures. The cold-start problem, the inability to detect novel attack types before sufficient labeled examples are available for training, remains largely unaddressed in the reviewed literature despite its critical operational significance. Class imbalance mitigation techniques including SMOTE, cost-sensitive learning, and anomaly-based formulations are unevenly applied, with fewer than 40% of the reviewed intrusion detection studies reporting explicit strategies for handling the extreme class skew characteristic of operational traffic.

5. DISCUSSION

5.1 SYNTHESIS OF FINDINGS AND EMERGING TRENDS

The meta-analysis presented in this paper reveals several clear trends in the evolution of ML and DL applications in cybersecurity over the review period. First, there has been a progressive shift from classical ML algorithms toward deep learning architectures, with LSTM and CNN-based models dominating recent

publications and consistently outperforming Random Forest and SVM baselines on sufficiently large datasets. Second, the increasing availability of diverse benchmark datasets including CICIDS2017, CIC-DDoS2019, and EMBER has enabled more rigorous comparative evaluation, although the field continues to lack a standardized evaluation protocol equivalent to those established in computer vision or natural language processing. Third, the emergence of transformer-based architectures and graph neural networks represents a frontier of research activity with significant untapped potential for cybersecurity applications where relational and sequential structure is present. The meta-analytical pooled estimates indicate that deep learning models achieve median F1-scores approximately 3-7 percentage points higher than classical ML across the primary application domains, with the advantage most pronounced for encrypted traffic classification and raw binary malware analysis where end-to-end feature learning provides the greatest benefits. However, this performance advantage is partially offset by significantly higher computational training costs, reduced interpretability, and greater sensitivity to adversarial manipulation.

5.2 FUTURE RESEARCH DIRECTIONS

Several high-priority research directions emerge from the critical analysis presented in this paper. Federated learning for privacy-preserving collaborative threat intelligence represents a particularly promising area, enabling organizations to jointly train detection models without sharing sensitive network traffic data, addressing both data scarcity and privacy compliance challenges. Explainable AI integration into cybersecurity ML pipelines requires concerted methodological development, with a focus on developing domain-specific interpretability frameworks that align model explanations with the analytical workflows of security practitioners. The development of adversarially robust evaluation benchmarks that explicitly incorporate adaptive adversary models is essential for producing research results with greater validity for operational deployments. Continual learning approaches that enable deployed models to adapt to distributional shift without catastrophic forgetting represent an important practical capability absent from the vast majority of reviewed works. Finally, the integration of causal inference methods into cybersecurity ML represents an underexplored avenue for improving both the generalizability of learned representations and the quality of explanations provided to human analysts, addressing the fundamental limitation that correlation-based models are inherently vulnerable to distribution shift and adversarial manipulation.

6. CONCLUSION

This paper has presented a systematic review and meta-analysis of machine learning and deep learning applications in cybersecurity, synthesizing findings from 232 peer-reviewed studies published between 2013 and 2024. The review demonstrates that ML and DL techniques have achieved remarkable performance improvements across intrusion detection, malware classification, phishing detection, and related domains, with deep learning architectures offering the strongest results on tasks involving raw data inputs and complex pattern structures. However, the meta-analysis also reveals persistent methodological weaknesses including benchmark saturation, class imbalance neglect, adversarial vulnerability, and the near-total absence of real-world deployment evaluations. The critical analysis establishes that much of the reported performance improvement

reflects methodological refinements and favorable experimental conditions rather than fundamental advances in detection capability. Future progress requires the cybersecurity ML community to adopt more rigorous evaluation standards, develop adversarially valid benchmarks, and prioritize explainability and operational deployability alongside raw accuracy metrics. The convergence of federated learning, continual learning, explainable AI, and causal inference with cybersecurity applications offers a productive research agenda for the next decade. The findings of this review provide a consolidated reference for researchers and practitioners seeking to navigate the rapidly evolving landscape of intelligent cybersecurity technologies.

REFERENCES

[1] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in Proc. IEEE Symp. Comput. Intell. Security Defense Appl., Ottawa, Canada, 2009, pp. 1–6.

[2] H.-J. Liao and C. H. Richard Lin and Y.-C. Lin and K.-Y. Tung, "Intrusion detection system: A comprehensive review," J. Netw. Comput. Appl., vol. 36, no. 1, pp. 16–24, 2013.

[3] S. Mukkamala, G. Janoski, and A. Sung, "Intrusion detection using neural networks and support vector machines," in Proc. IEEE Int. Joint Conf. Neural Netw., Honolulu, HI, USA, 2002, vol. 2, pp. 1702–1707.

[4] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[5] M. Panda and M. R. Patra, "Network intrusion detection using naive Bayes and random forest," Int. J. Recent Trends Eng., vol. 1, no. 1, pp. 420–422, 2009.

[6] N. Farnaaz and M. A. Jabbar, "Random forest modeling for network intrusion detection system," Procedia Comput. Sci., vol. 89, pp. 213–217, 2016.

[7] S. S. Dhaliwal, A. A. Nahid, and R. Abbas, "Effective intrusion detection system using XGBoost," Information, vol. 9, no. 7, p. 149, 2018.

[8] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in Proc. Int. Conf. Inf. Syst. Security Privacy, 2018, pp. 108–116.

[9] R. Panigrahi and S. Borah, "A detailed analysis of CICIDS2017 dataset for designing intrusion detection systems," Int. J. Eng. Technol., vol. 7, no. 3.24, pp. 479–482, 2018.

[10] M. Belouch, S. El Hadaj, and M. Idhammad, "A two-stage classifier approach using RepTree algorithm for network intrusion detection," Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 6, 2017.

[11] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in Proc. 9th EAI Int. Conf. Bio-inspired Inf. Commun. Technol., 2016, pp. 21–26.

[12] Y. Li, R. Ma, and R. Jiao, "A hybrid malicious code detection method based on deep learning," Int. J. Security Appl., vol. 9, no. 5, pp. 205–216, 2015.

[13] W. Wang, M. Zhu, J. Wang, X. Zeng, and Z. Yang, "End-to-end encrypted traffic classification with one-dimensional convolution neural networks," in Proc. IEEE Int. Conf. Intell. Security Inf., 2017, pp. 43–48.

[14] G. Andresini, F. Pendlebury, F. Pierazzi, C. Loglisci, A. Appice, and L. Cavallaro, "INSOMNIA: Towards concept-drift robustness in network intrusion detection," in Proc. ACM Workshop Artif. Intell. Security, 2021, pp. 111–122.

[15] W. W. Lo, S. Layeghy, M. Sarhan, M. Gallagher, and M. Portmann, "E-GraphSAGE: A graph neural network based intrusion detection system for IoT," in Proc. IEEE/IFIP Netw. Operations Manage. Symp., 2022, pp. 1–9.

[16] C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," IEEE Access, vol. 5, pp. 21954–21961, 2017.

[17] J. Zhao, X. Shetty, J. W. Pan, C. Kamhoua, and K. Kwiat, "Transfer learning for detecting unknown network attacks," EURASIP J. Inf. Security, vol. 2019, no. 1, pp. 1–13, 2019.

[18] M. G. Schultz, E. Eskin, E. Zadok, and S. J. Stolfo, "Data mining methods for detection of new malicious executables," in Proc. IEEE Symp. Security Privacy, Oakland, CA, USA, 2001, pp. 38–49.

[19] S. Forrest, S. A. Hofmeyr, A. Somayaji, and T. A. Longstaff, "A sense of self for Unix processes," in Proc. IEEE Symp. Security Privacy, 1996, pp. 120–128.

[20] K. Rieck, P. Trinius, C. Willems, and T. Holz, "Automatic analysis of malware behavior using machine learning," J. Comput. Security, vol. 19, no. 4, pp. 639–668, 2011.

[21] L. Nataraj, S. Karthikeyan, G. Jacob, and B. S. Manjunath, "Malware images: Visualization and automatic classification," in Proc. 8th Int. Symp. Visualization Cyber Security, 2011, pp. 1–7.

[22] D. Vasan, M. Alazab, S. Wassan, B. Safaei, and Q. Zheng, "Image-based malware classification using ensemble of CNN architectures (IMCEC)," Comput. Security, vol. 92, p. 101748, 2020.

[23] O. David and N. S. Netanyahu, "DeepSign: Deep learning for automatic malware signature generation and classification," in Proc. Int. Joint Conf. Neural Netw., 2015, pp. 1–8.

[24] W. Hu and Y. Tan, "Generating adversarial malware examples for black-box attacks based on GAN," arXiv preprint arXiv:1702.05983, 2017.

[25] D. Gibert, C. Mateu, and J. Planes, "The rise of machine learning for detection and classification of malware: Research developments, trends, and challenges," J. Netw. Comput. Appl., vol. 153, p. 102526, 2020.

[26] M. Khonji, Y. Iraqi, and A. Jones, "Phishing detection: A literature survey," IEEE Commun. Surveys Tuts., vol. 15, no. 4, pp. 2091–2121, 2013.

[27] R. M. Mohammad, F. Thabtah, and L. McCluskey, "Predicting phishing websites based on self-structuring neural network," Expert Syst. Appl., vol. 40, no. 13, pp. 4966–4971, 2013.

[28] P. Yang, G. Zhao, and P. Zeng, "Phishing website detection based on multidimensional features driven by deep learning," IEEE Access, vol. 7, pp. 15196–15209, 2019.

[29] W. Lo, J. Griffiths, and Z. Xu, "Applying social network analysis to the detection of network intrusions," in Proc. Int. Conf. Comput. Intell. Security, 2010, pp. 1–6.

[30] I. Fette, N. Sadeh, and A. Tomasic, "Learning to detect phishing emails," in Proc. 16th Int. Conf. World Wide Web, 2007, pp. 649–656.