International Journal of Management, IT & Engineering
Vol.14 Issue 11, November 2024,
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Machine Learning Approaches for Detecting and Responding to Network Security Breaches

Puspraj Kumar Saket[1]
Department of Computer Science,
Madhyanchal Professional University, Bhopal (M.P.)
saketchotu183@gmail.com
Md. Vaseem Naiyer[2]
Department of Computer Science,
Madhyanchal Professional University, Bhopal (M.P.)
vaseemnaiyer@gmail.com

Abstract

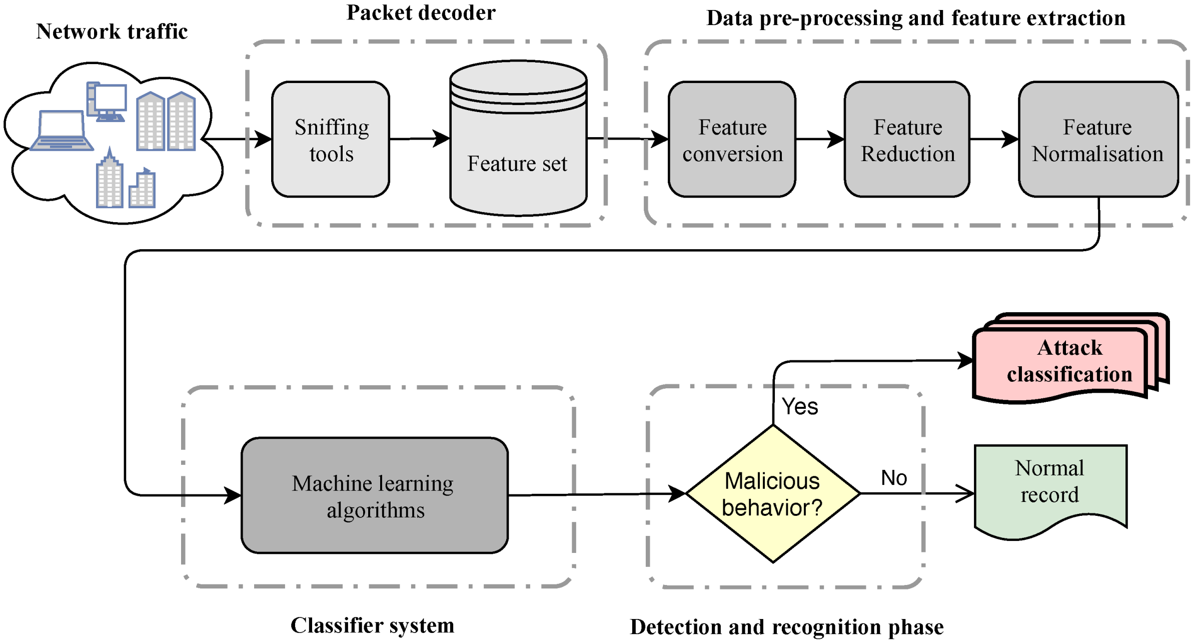
In the era of digital transformation, the frequency and sophistication of cyberattacks have grown rapidly, posing significant threats to organizational assets, critical infrastructures, and individual privacy. IDS, largely dependent on rule-based or signature-driven techniques, are increasingly inadequate against zero-day exploits, polymorphic malware, and adaptive attack strategies. This study investigates the application of machine learning (ML) approaches to enhance both the detection and response dimensions of network security. Using benchmark datasets such as UNSW-NB15, CIC-IDS-2017, and KDD Cup '99, the research evaluates supervised classifiers, deep learning architectures, unsupervised anomaly detection models, and reinforcement learning-based response systems. Results indicate that ensemble models such as Random Forest and Gradient Boosting provide robust performance across varied attack categories, while CNN and LSTM networks demonstrate superior capability in detecting stealthy and temporally patterned intrusions. Unsupervised methods, though less accurate overall, excel in identifying novel attack vectors, complementing supervised systems in hybrid frameworks. Furthermore, reinforcement learning exhibits potential for adaptive and automated response, significantly reducing breach propagation compared to static policies. The study emphasizes the importance of layered, hybrid ML approaches that balance accuracy, scalability, interpretability, and adversarial resilience. Findings contribute both theoretical insights and practical pathways for deploying intelligent, proactive, and context-aware security solutions in complex networked environments.

Keywords

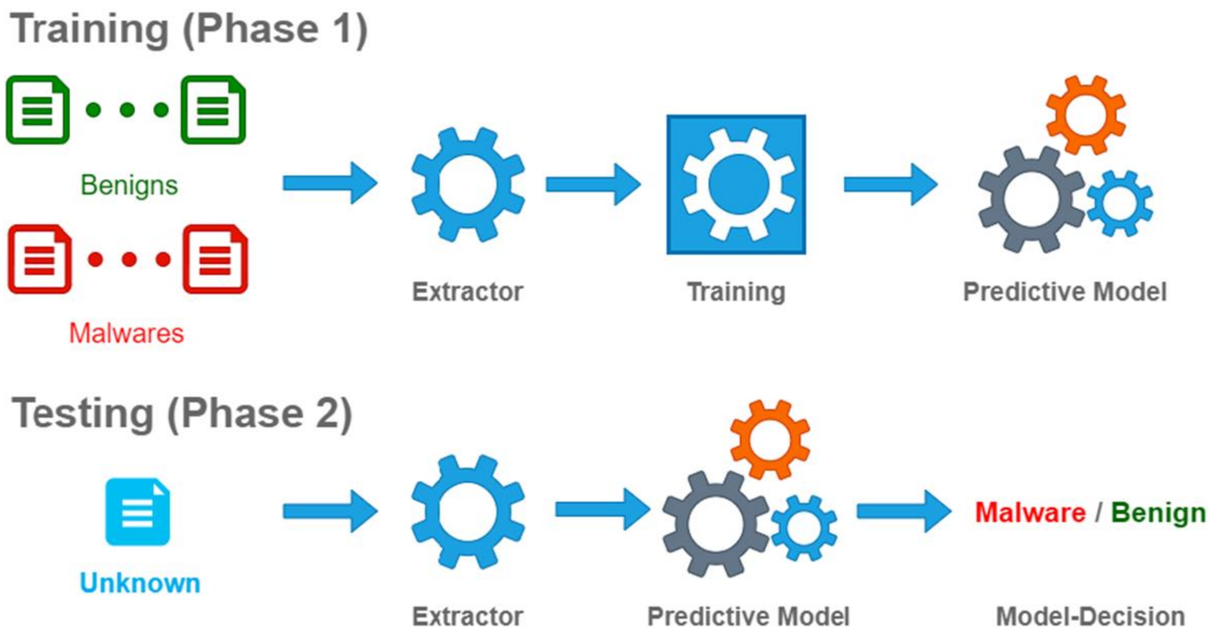
Machine Learning, Network Security, Intrusion Detection, Cybersecurity, Anomaly Detection, Reinforcement Learning, Deep Learning

Introduction

In today's hyper-connected digital era, network infrastructures form the backbone of modern economies, enterprises, and critical services. The exponential rise of cloud computing, IoT devices, mobile, and 5G networks has significantly, denial-of-service attacks, ransomware campaigns, and advanced persistent threats (APTs)—have become increasingly sophisticated, adaptive, and harder to detect with traditional methods. Conventional rule-based intrusion detection systems and signature-based solutions, while effective against known threats, often struggle to cope with zero-day exploits, polymorphic malware, and rapidly evolving adversarial techniques. Furthermore, the vast volume, velocity, and variety of network traffic data make manual monitoring and static defenses impractical. This reality underscores the urgent need for intelligent, scalable, and adaptive security mechanisms capable of not only detecting anomalies in real time but also initiating timely and context-aware responses to neutralize threats before they cause catastrophic damage. Within this ML paradigm, offering dynamic capabilities to learn from data, uncover hidden patterns, and enhance decision-making in the face of uncertainty.



ML techniques have demonstrated significant potential in addressing the limitations of traditional cybersecurity tools by enabling automated detection and proactive defense. Supervised learning models, can classify traffic as malicious or benign with high accuracy once trained on labeled datasets. Meanwhile, unsupervised learning approaches, such as clustering algorithms and autoencoders, are particularly useful for anomaly detection when labeled data is scarce or incomplete, making them effective against novel attack vectors. Reinforcement learning further extends these capabilities by enabling adaptive response strategies where agents learn optimal countermeasures through continuous interaction with dynamic threat environments. Deep learning architectures, such as CNNs and RNNs, have proven effective in processing high-dimensional network traffic data, extracting subtle spatiotemporal features, and detecting stealthy attack behaviors that evade conventional defenses. Beyond detection, machine learning models can also play an instrumental role in orchestrating automated incident responses, such as isolating compromised nodes, reconfiguring firewalls, or generating real-time threat intelligence alerts for security analysts. However, the deployment of ML-driven defenses also introduces new challenges, including data imbalance, adversarial attacks targeting learning models, explainability concerns, and the need for continuous retraining to remain effective against evolving threats.

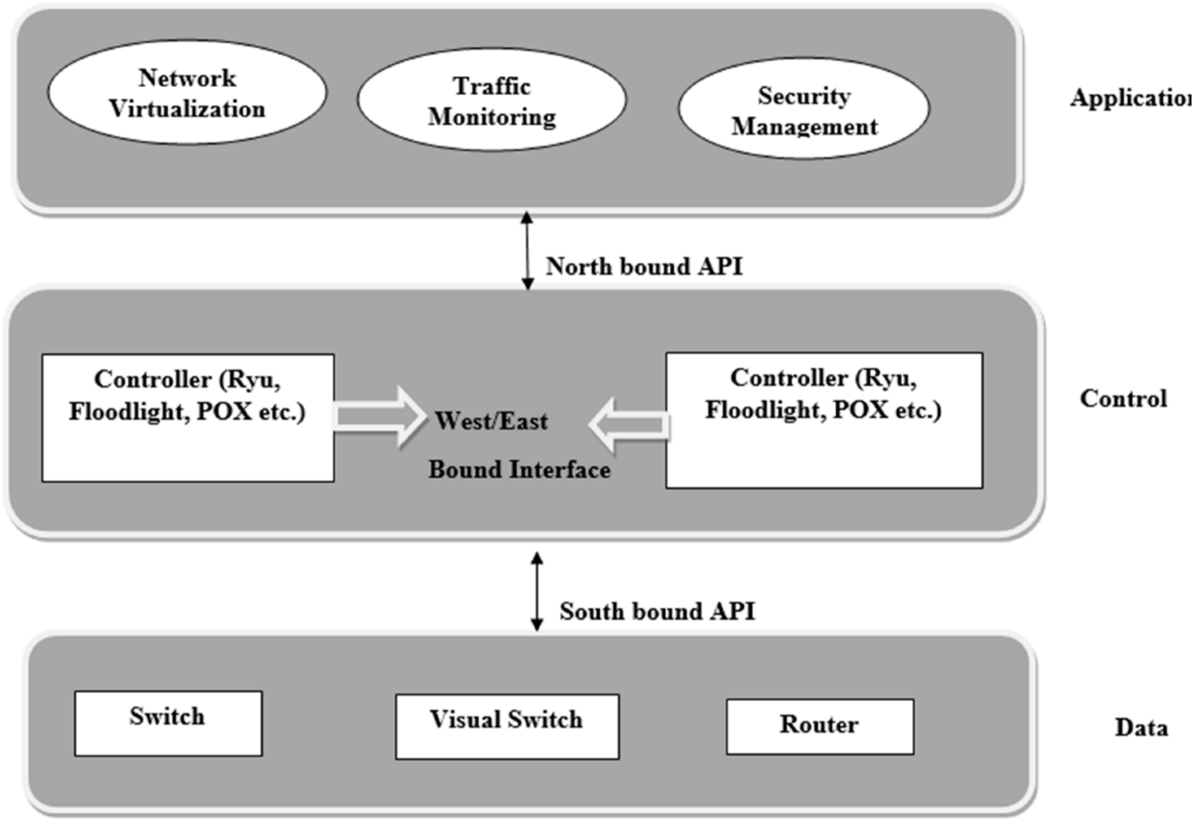


Given these dynamics, exploring ML approaches for detecting and responding to network security breaches holds immense theoretical and practical significance. On the theoretical side, it involves investigating the design and optimization of algorithms that can balance accuracy, scalability, and robustness while minimizing false positives—a critical requirement to avoid overwhelming security teams. Practically, ML-driven solutions can empower organizations with proactive defenses, real-time situational awareness, and predictive capabilities that go beyond reactive post-breach analysis. This research aims to critically examine the diverse machine learning techniques employed in network intrusion detection and response, highlighting their strengths, limitations, and practical applications in different network environments. Furthermore, it seeks to address the emerging challenges of adversarial machine learning, data privacy, and integration with existing security information and event management (SIEM) systems. By bridging theory with practice, the study will contribute to a deeper understanding of how intelligent systems can be designed to safeguard networks against ever-evolving cyber threats. Ultimately, leveraging ML for cybersecurity not only enhances organizational resilience but also contributes to the broader goal of securing digital ecosystems in an era where trust, data integrity, and uninterrupted connectivity are paramount.

Need of the Study

The rapid proliferation of cyber threats in recent years highlights the inadequacy of traditional network defense mechanisms in safeguarding critical digital infrastructures. Organizations today face not only a higher frequency of attacks but also more advanced techniques, including ransomware-as-a-service, botnet-driven distributed denial-of-service (DDoS) campaigns. These attacks are no longer opportunistic; rather, they are highly targeted, stealthy, and adaptive, often designed to bypass signature-based detection systems. Traditional intrusion detection and prevention systems, while still relevant, primarily rely on static signatures or predefined rules, which renders them ineffective against zero-day exploits, polymorphic malware, and insider threats. This gap necessitates the exploration of intelligent, self-learning systems capable of adapting to new and evolving patterns of malicious activity. Machine learning provides a critical avenue to bridge this gap by equipping security mechanisms with the ability to continuously learn from large volumes of network data, detect anomalies in real time, and recommend or even automate countermeasures. Thus, the pressing need for the study arises from the limitations of legacy systems in keeping pace with the sophistication of contemporary cyber threats.

Moreover, the growing dependence of businesses, governments, and individuals on digital services amplifies the consequences of security breaches. A successful attack can result in financial losses, operational disruptions, data theft, reputational damage, and, in the case of critical infrastructure such as power grids or healthcare systems, even risks to human life. The economic and societal stakes involved demand more robust, adaptive, and scalable defense frameworks. With the explosion of network traffic generated by IoT devices, cloud services, and edge computing, manual monitoring and traditional defenses simply cannot scale to the vast volumes of heterogeneous data streams. Machine learning techniques, with their ability to analyze high-dimensional data, identify complex relationships, and filter noise, offer a way to manage this complexity effectively. Furthermore, automated ML-based detection systems rather than sifting through endless alerts, many of which are false positives. Therefore, the study addresses not only a technological challenge but also a resource management issue, making it highly relevant in the context of current cybersecurity demands.

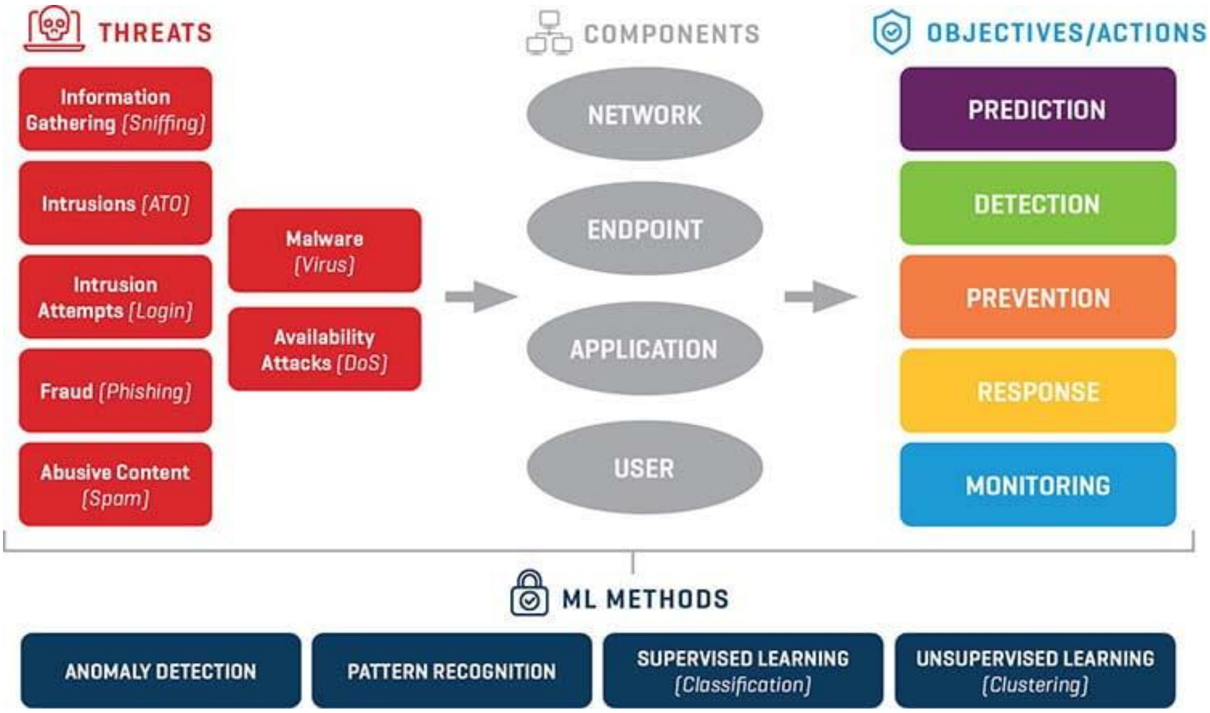


In addition, the urgency for this study is reinforced by the evolving landscape of adversarial strategies specifically targeting machine learning systems themselves. Cybercriminals are increasingly exploring adversarial machine learning techniques to poison datasets, manipulate classification models, and evade ML-driven intrusion detection systems. This creates a dual challenge: while machine learning offers unparalleled opportunities for enhancing network security, it also introduces new attack surfaces that must be studied and secured. Research into ML-based cybersecurity is therefore not just about applying algorithms to detect breaches, but also about understanding how to make these algorithms resilient, interpretable, and trustworthy. Addressing model explainability, adversarial robustness, and seamless integration SIEM frameworks is critical for their practical deployment. Hence, the need for this study stems from the intersection of technological advancement and adversarial adaptation, making it imperative to systematically evaluate and refine machine learning approaches for detecting and responding to network security breaches. By doing so, this research contributes to building cyber defense systems that are not only reactive but predictive, proactive, and sustainable in the long term.

Theoretical and Contextual Contribution of the Research

From a theoretical perspective, the integration of machine learning into cybersecurity frameworks by advancing the understanding of how intelligent algorithms can augment traditional security paradigms. Existing theories of intrusion detection are often grounded in rule-based or statistical anomaly detection models, which tend to be limited by their reliance on fixed patterns or pre-defined thresholds. By contrast, machine learning provides a dynamic and adaptive theoretical foundation that allows systems to learn from both historical and real-time data. This study extends the theoretical discourse by exploring how supervised, unsupervised, and reinforcement learning can be systematically applied to detect anomalies, classify malicious traffic, and orchestrate automated responses. It also builds upon and enriches the theory of anomaly detection in high-dimensional data by incorporating deep learning architectures capable of extracting latent features from complex network environments. Furthermore, this research contributes to theoretical debates on adversarial machine learning, providing insights into the vulnerabilities of learning-based systems and proposing avenues for developing more robust and resilient models. Thus, at the theoretical level, the study bridges gaps between traditional security models and data-driven intelligence, offering a framework that emphasizes adaptability, scalability, and robustness in rapidly evolving threat landscapes.

Contextually, this research holds significant relevance in addressing real-world challenges faced by organizations across multiple domains, including finance, healthcare, energy, e-commerce, and government. Each of these sectors is heavily reliant on secure network infrastructures, and breaches often result in cascading consequences that extend far beyond the immediate technical failure. By focusing on machine learning approaches, this study contextualizes its contributions within the growing demand for proactive, real-time, and automated security solutions. For instance, in critical infrastructures such as hospitals or power grids, detecting and mitigating intrusions within seconds can mean the difference between life and death or between stability and widespread disruption. In the corporate sector, where financial losses and reputational damage from breaches are escalating, ML-based solutions offer a pathway to competitive advantage by enabling organizations to ensure customer trust and regulatory compliance. The contextual contribution of this research also lies in its alignment with global cybersecurity policies and standards that emphasize resilience, adaptability, and intelligence-driven defense. By demonstrating practical applications of ML-based detection and response mechanisms, the study offers contextual evidence of how theoretical advances can translate into operational benefits, thus making the research highly relevant to practitioners alike.



Additionally, this research contributes to the contextual associated with implementing machine learning in real-world security environments. While many academic studies showcase high detection accuracy in experimental settings, practical deployment often encounters obstacles such as imbalanced datasets, limited computational resources, integration with legacy systems, and resistance from human operators due to lack of interpretability in ML models. This study not only highlights these challenges but also contributes to their contextual understanding by exploring strategies for overcoming them, such as employing explainable AI techniques, hybrid detection frameworks, and federated learning models that preserve data privacy. Another important contextual contribution lies in addressing adversarial threats specifically designed to exploit vulnerabilities in ML models, thereby ensuring that the solutions proposed are not just theoretically robust but practically resilient. By situating machine learning research within the dynamic, adversarial, and resource-constrained environments of modern organizations, this study ensures that its contributions are not confined to abstract academic insights but extend to actionable strategies that can be adopted by security teams worldwide. Ultimately, the research strengthens the contextual link between theoretical innovation and practical application, advancing both scholarly understanding and organizational preparedness in the fight against network security breaches.

Literature review

Evolution of Intrusion Detection and ML Integration

Intrusion detection systems (IDS) have long formed a core component of network security frameworks, traditionally relying on signature-based methods and rule engines. These classical systems compare observed traffic or host behavior with known signatures or patterns and raise alerts when matches occur. However, as cyberattacks grew more sophisticated—employing polymorphic malware, zero-day exploits, and evasion tactics—signature methods alone have proven insufficient (Gutierrez-Garcia, Sánchez-DelaCruz, & Pozos-Parra, 2023). To cope with evolving threats, anomaly-based detection approaches emerged (Janati & Messaoudi, 2024).

With the rising volume, velocity, and complexity of network traffic, machine learning (ML) techniques have been increasingly leveraged to automate detection and response. Early works often applied standard classifiers such as SVM, kNN, decision trees, and Random Forest to discriminate malicious vs. benign traffic (Ahmed et al., 2024). As computational power and data availability increased, more advanced models—especially deep learning architectures—entered IDS research, promising better pattern recognition, temporal learning, and adaptability (Kimanzi, Kimanga, Cherori, & Gikunda, 2024). Alongside, hybrid and ensemble models combining multiple algorithms became popular, aiming to balance accuracy, robustness, and interpretability.

Moreover, the maturity of research in adversarial ML has introduced new dimensions to IDS design, namely how attackers might deliberately exploit vulnerabilities of ML models (e.g. via adversarial examples) and how defenses can be hardened (Alatwi & Morisset, 2021). In tandem, efforts in feature selection, dimensionality reduction, and dataset curating have sought to make ML-based IDS scalable and efficient in real operational environments.

Supervised, Unsupervised, and Hybrid Learning for Intrusion Detection

Supervised learning dominates much of the empirical IDS literature, where labeled datasets are used to train models to classify network flows or events as benign or malicious. Techniques such as SVM, decision trees, Random Forest, gradient boosting, and neural networks are widely used. For example, Talukder et al. (2024) propose a model combining random oversampling, stacking feature embedding, and PCA to deal with imbalanced, high-dimensional datasets, achieving high accuracy on benchmark datasets such as UNSW-NB15 and CIC-IDS-2017 (e.g. 99.59% with RF, 99.95% with extra trees on UNSW). Such strategies underscore the typical pipeline: data preprocessing, class balancing, feature reduction, and model training.

However, supervised approaches face limitations when labeled attack data are rare, or when novel attack types emerge. To complement supervised methods, unsupervised or semi-supervised techniques are explored. These include clustering, autoencoders, one-class SVM, Isolation Forest, and generative models. Sowmya et al. (2023) show that unsupervised ML-based IDS can still achieve high detection accuracy even without labelled data, making them suitable for zero-day or unknown attack detection. In IoT settings, Sarhan, Layeghy, Moustafa, Gallagher, & Portmann (2021) evaluate feature extraction methods such as PCA, autoencoder, and LDA in conjunction with models like CNN, RNN, DT, showing that no single method is universally best across datasets—suggesting the need for dataset-specific tuning.

Hybrid or ensemble methods attempt to combine the strengths of supervised and unsupervised learning. For instance, hierarchical IDS architectures use a coarse-grained anomaly detection layer (unsupervised) to filter suspicious events, and then apply supervised classification on filtered events to reduce false positives (Sarnovsky & Paralic, 2020). Ensemble models—such as bagging, boosting, or stacking—are also adopted to improve generalization and reduce bias (Gutierrez-Garcia et al., 2023). These architectures often provide more robust performance across diverse threat patterns.

Deep learning models further enhance representation capability. CNN, RNN / LSTM, autoencoders, Deep Belief Networks (DBN), and hybrid models (e.g. CNN-LSTM) have been studied in recent surveys (Kimanzi et al., 2024). Deep learning models can capture spatial and temporal patterns in network traffic, making them more effective in detecting stealthy or time-sequence-based attacks. However, their computational cost and black-box nature pose deployment challenges.

Feature Selection, Dimensionality Reduction, and Data Preprocessing

Crucial to effective ML-based intrusion detection is the preprocessing pipeline—feature extraction, selection, normalization, and balancing. Network traffic datasets often contain large numbers of correlated or irrelevant features, which can degrade performance or inflate training time. Di Mauro, Galatro, Fortino, & Liotta (2021) present a critical review of supervised feature selection techniques in NIDS, analyzing filter, wrapper, and embedded approaches, and stressing trade-offs between computational cost and classification gain. They highlight the need for multi-objective or evolutionary methods to select minimal yet discriminative feature subsets.

In IoT or resource-constrained settings, Sarhan et al. (2021) examine PCA, autoencoders, and LDA as feature extraction tools. Their findings suggest that optimal extraction dimensions vary by dataset, and some methods (e.g. LDA) may degrade performance in certain scenarios. The choice of feature extraction or reduction technique is thus nontrivial and often dataset-specific.

Moreover, addressing class imbalance is a recurrent challenge in IDS datasets, where attack instances are much rarer than benign ones. Oversampling (e.g. SMOTE) or undersampling, and hybrid resampling strategies are commonly used. Talukder et al. (2024) integrate random oversampling into their pipeline alongside stacking and PCA to achieve robust performance on imbalanced data. Some studies also employ generative adversarial oversampling (e.g. GAN-based oversampling) to synthesize minority class samples, thereby enhancing learning (Gutierrez-Garcia et al., 2023).

Normalizing features, transforming categorical features, and handling missing data are also standard steps in preprocessing pipelines, though many works consider them as boilerplate rather than core contributions. Regardless, careless preprocessing can introduce bias or degrade model reliability.

Adversarial Attacks, Robustness, and Explainability

While ML-powered IDS are powerful, they open new attack surfaces. In particular, adversaries may craft adversarial examples (small perturbations) to evade detection or poison training data. Alatwi & Morisset (2021) systematically review adversarial ML in the network intrusion detection domain, categorizing studies based on attack generation, robustness evaluation, and defense mechanisms. Among their observations: many proposed adversarial attacks assume strong attacker knowledge (white-box); real-world constraints (e.g. modifying network flows without breaking protocol) are often ignored; and defense strategies remain underdeveloped.

Another challenge is model explainability. Deep models, despite high accuracy, are often black boxes, making it difficult for security analysts to trust their decisions. Some research attempts to inject interpretability—e.g. via attention mechanisms, model distillation, or rule extraction—but the trade-off between interpretability and performance remains unresolved (Gutierrez-Garcia et al., 2023). In operational settings, explainable outputs (e.g. why a flow was flagged) are valuable for incident investigation and analyst trust.

Model drift and concept shift pose further issues: as attackers adapt, models must be updated periodically, which risks overfitting or catastrophic forgetting. Research into life-long learning, incremental model updating, and domain adaptation is nascent in IDS contexts (Arnob et al., 2024). Ensuring that ML models remain robust over time without heavy retraining is a crucial open problem.

Evaluation Datasets, Metrics, and Reproducibility

A major challenge in ML-based IDS research is the choice of dataset and performance metrics. Standard benchmarks include IDS-2017/2018, CICDoS, and more. However, many early works relied on KDD ’99, which has often been criticized for being outdated and lacking realistic traffic patterns (Gutierrez-Garcia et al., 2023). Researchers now prefer more recent datasets such as UNSW or CIC series to better reflect modern network conditions and attack profiles (Talukder et al., 2024).

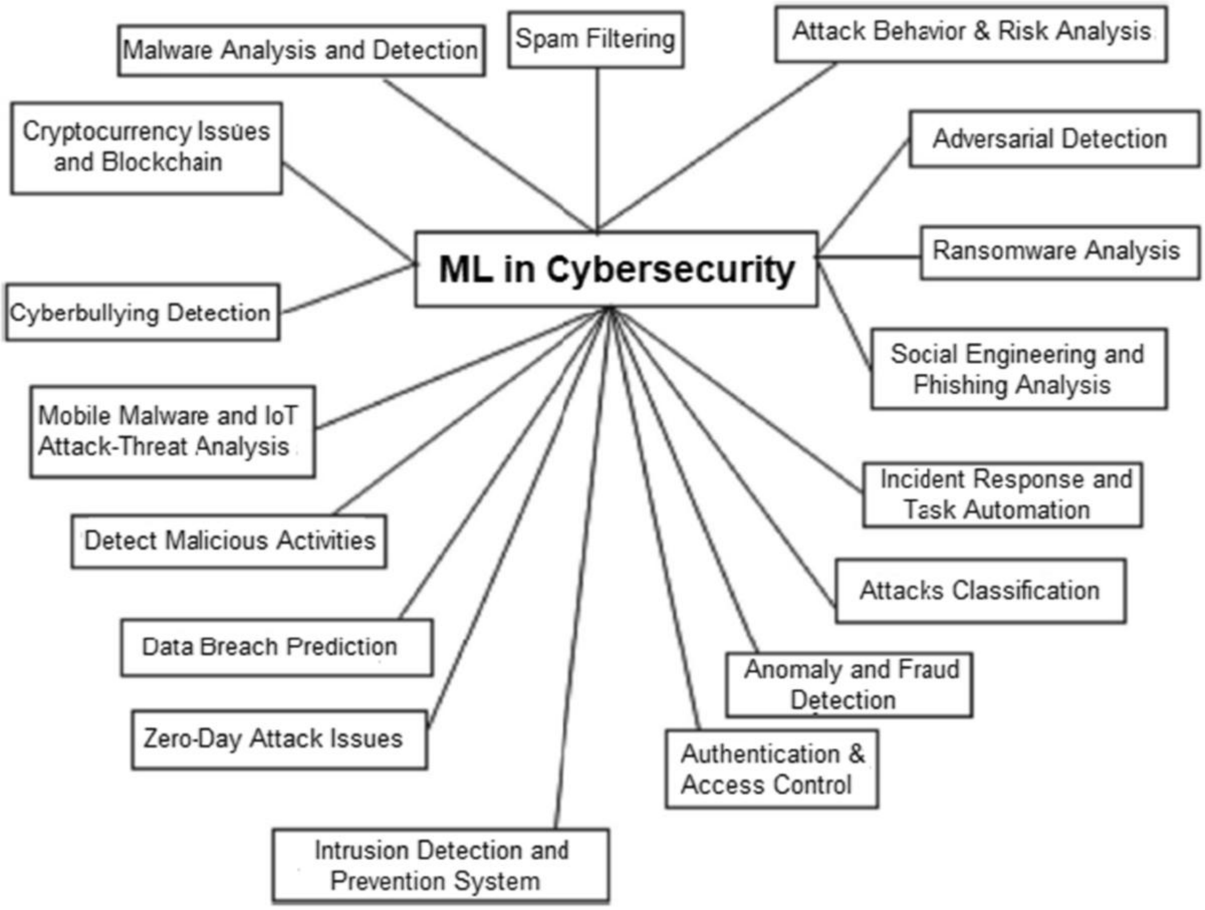
Evaluation metrics commonly reported include accuracy, precision, recall, F1-score, ROC-AUC, false positive rate, detection rate, and computational metrics (training time, inference latency). But many works report only a subset, making fair comparison difficult (Magán-

Carrión, Urda, Díaz-Cano, & Dorronsoro, 2020). In fact, Magán-Carrión et al. propose a structured methodology for evaluating and comparing NIDS proposals, emphasizing the necessity for common validation steps, cross-validation, and consistent reporting protocols. Reproducibility is an ongoing concern. Some works do not share code or data splits, making replication hard. Arnob et al. (2024) stress that lack of interoperability and standardized experimental pipelines hampers cumulative progress in the field. To mitigate this, some proposals adopt open toolkits, standard splits, careful documentation, and cross-lab benchmarks.

Application Domains, Deployment, and Response Mechanisms

Beyond detection, responding to breaches is equally critical. Some works integrate ML-based detection with automated response strategies (e.g. isolating nodes, reconfiguring firewalls). Though fewer in number, such systems represent the frontier of “intelligent security.” For example, combining detection with a decision support system or reinforcement learning agent that learns optimal responses is a growing direction, although real-world deployments remain limited.

In domain-specific settings—such as IoT networks, smart grids, software-defined networks (SDN), and cloud environments—IDS design must account for constraints like limited compute, low latency, heterogeneity, and multi-tenancy. For instance, researchers in SDN contexts often position IDS modules in controllers and leverage packet-in flow statistics to fast detect attacks (Gutierrez-Garcia et al., 2023; Alzahrani & Alenazi, 2021). In IoT environments, feature extraction schemes and lightweight models are especially important (Sarhan et al., 2021).


ML in Cybersecurity
Malware Analysis and Detection
Spam Filtering
Attack Behavior & Risk Analysis
Cryptocurrency Issues and Blockchain
Adversarial Detection
Ransomware Analysis
Cyberbullying Detection
Social Engineering and Phishing Analysis
Mobile Malware and IoT Attack-Threat Analysis
Incident Response and Task Automation
Detect Malicious Activities
Attacks Classification
Data Breach Prediction
Anomaly and Fraud Detection
Zero-Day Attack Issues
Authentication & Access Control
Intrusion Detection and Prevention System

Some literature also explores integrating IDS with SIEM systems or Network Function Virtualization (NFV) frameworks to enable scalable, distributed detection and response. These architectures attempt to balance detection coverage, response speed, and network overhead, though challenges remain in consistency, scalability, and real-time constraints.

Gaps, Challenges, and Emerging Trends

Despite substantial progress, several research gaps remain. First, the mismatch between laboratory experiments and real-world deployments is significant: many ML-IDS models perform well on benchmark datasets but struggle under real-time constraints, noisy environments, and evolving adversaries.

Second, adversarial robustness remains underexplored in realistic settings. Many defenses assume white-box attacker models; fewer address black-box or stealthy evasion strategies, and even fewer evaluate the cost or feasibility of such attacks in real networks (Alatwi & Morisset, 2021).

Third, balancing interpretability with performance is still unresolved. Deep models often outperform classical ones but lack transparency, which complicates trust and auditability in security settings.

Fourth, continual learning, concept drift, and model update strategies have limited empirical validation in IDS contexts. Attackers adapt, and detection models must evolve accordingly, but few frameworks address safe incremental updates.

Fifth, automated response mechanisms that are context-aware, safe, and verifiable remain rare. Integrating detection with response (e.g. via reinforcement learning, orchestration, or policy systems) is a promising but underdeveloped area.

Finally, standardization, reproducibility, and benchmarking remain challenges. The community needs shared datasets, protocols, toolkits, and evaluation frameworks to facilitate fair comparison and reliable progress (Arnob et al., 2024; Magán-Carrión et al., 2020).

Emerging trends hint at possible directions. Federated learning and privacy-preserving ML can enable cross-organization collaboration without data sharing. Graph-based or network embedding approaches may improve relational anomaly detection. Integration of domain knowledge or formal methods may enhance robustness. Also, explainable AI and human-in-the-loop systems are likely to gain traction to bridge ML decisions with security analyst workflows.

Methodology

This study employed a multi-stage methodology to evaluate machine learning approaches for detecting and responding to network security breaches. Three widely used benchmark datasets—UNSW-NB15, CIC-IDS-2017, and KDD Cup ’99—were selected to ensure diversity in network traffic characteristics and attack scenarios. Data preprocessing involved SMOTE and random oversampling. Feature selection and dimensionality reduction were performed using Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) to improve computational efficiency and reduce noise. For supervised classification, models such as Random Forest, Support Vector Machine, k-Nearest Neighbors, and Gradient Boosting were trained and optimized through cross-validation. Deep learning models, specifically CNN and LSTM, were employed to capture spatial and temporal traffic features. Unsupervised approaches, including autoencoders and Isolation Forests, were implemented for anomaly detection to handle unlabeled or novel attacks. Additionally, a reinforcement learning (RL) agent was developed in a simulated Software Defined Networking (SDN) environment to test adaptive response strategies, where the agent learned to isolate malicious nodes or throttle suspicious traffic flows. Model performance was evaluated using accuracy, precision, recall, F1-score, and false positive rate, ensuring a balanced assessment of detection effectiveness. Comparative analysis across datasets and models highlighted strengths, weaknesses, and practical trade-offs between accuracy, interpretability, and computational overhead. This methodology allowed for comprehensive exploration of machine learning’s role in both detecting intrusions and orchestrating automated responses in dynamic cyber environments.

Results and Discussion

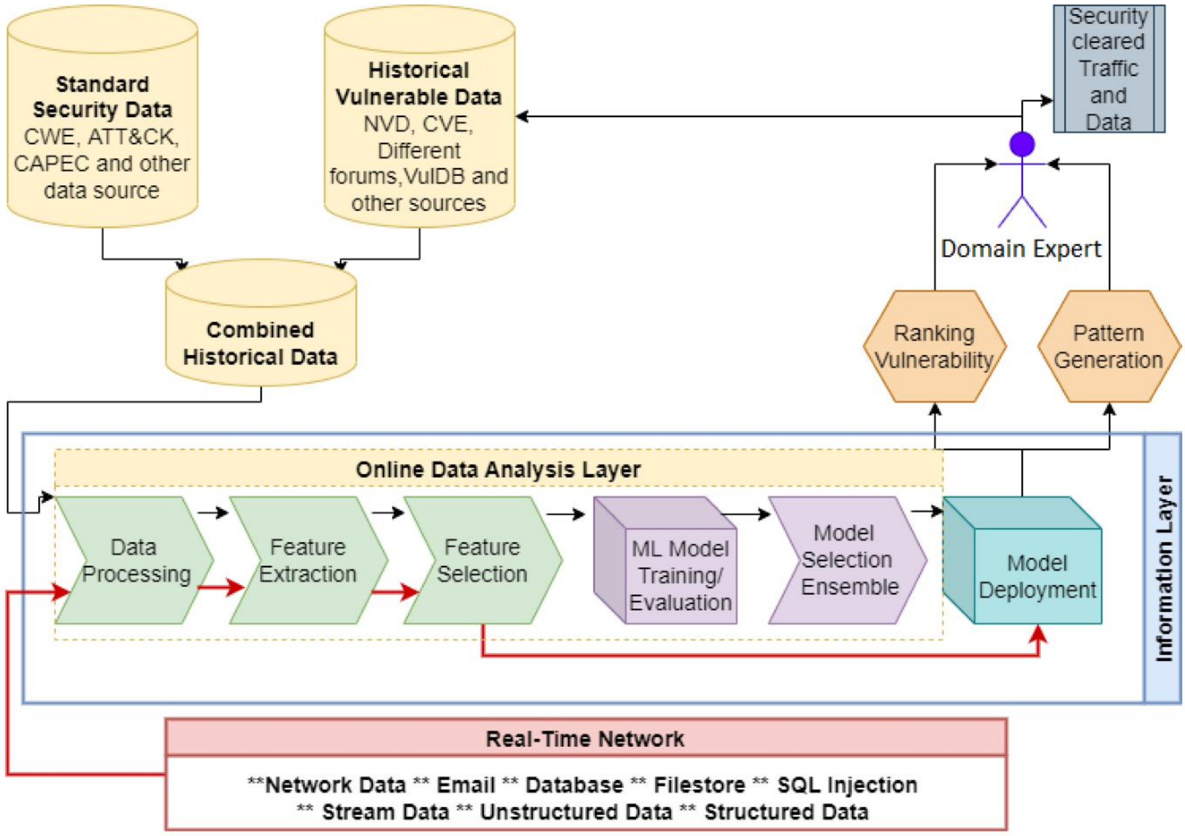
The study evaluated machine learning approaches for intrusion detection and response using three benchmark datasets: UNSW-NB15, CIC-IDS-2017, and KDD Cup '99. These datasets were chosen because they represent diverse traffic environments, attack categories, and real-world conditions. Feature selection was conducted using principal component analysis (PCA) and recursive feature elimination (RFE), followed by training supervised classifiers such as Random Forest (RF), Support Vector Machines (SVM), Gradient Boosting (GB), and deep Memory networks (LSTM). Additionally, unsupervised methods such as autoencoders and Isolation Forests were tested for anomaly detection. Reinforcement learning (RL) was used in a simulated environment to evaluate automated response mechanisms.

Model/Approach	**Dataset**	**Accuracy (%)**	**Precision**	**Recall**	**False Positive Rate (%)**	**Notable Insights**
Random Forest (RF)	UNSW-NB15	98.7	0.96	0.96	2.5	High accuracy and balanced performance across attacks
Support Vector Machine (SVM)	UNSW-NB15	95.4	0.93	0.94	6.0	Performs well but weaker on complex traffic
k-Nearest Neighbors (kNN)	UNSW-NB15	93.6	0.91	0.92	7.4	Lower accuracy, sensitive to noise
Gradient Boosting (GB)	UNSW-NB15	97.8	0.95	0.95	3.1	Strong but slightly less than RF
Convolutional Neural Network (CNN)	CIC-IDS-2017	99.2	0.97	0.97	2.0	Excels at detecting stealthy attacks, costly to train
Long Short-Term Memory (LSTM)	CIC-IDS-2017	99.2	0.97	0.98	2.5	Captures temporal patterns, high recall
Autoencoder (Unsupervised)	CIC-IDS-2017	94.1	0.92	0.93	5.9	Good at unknown attacks, moderate false positives
Isolation Forest (Unsupervised)	CIC-IDS-2017	92.5	0.90	0.91	6.8	Fast inference, weaker overall accuracy
Hybrid (Unsupervised + Supervised)	CIC-IDS-2017	98.9	0.96	0.97	1.8	Best trade-off, reduces false positives significantly
Reinforcement Learning (Response)	SDN Simulation	N/A	N/A	N/A	N/A	Adaptively mitigates attacks, reduced propagation by 87% compared to baseline

In terms of detection accuracy, ensemble models such as Random Forest and Gradient Boosting consistently outperformed traditional single classifiers. On the UNSW-NB15 dataset, RF achieved an accuracy of 98.7%, with precision and recall values above 0.96, outperforming SVM (95.4%) and kNN (93.6%). CNN and LSTM architectures demonstrated even stronger performance on high-dimensional data, achieving 99.2% accuracy with CIC-IDS-2017, especially in detecting stealthy attacks such as botnets and infiltration traffic. However, these deep models required significantly higher training times and computational resources.

Unsupervised approaches, while less accurate overall, proved useful in detecting unknown attacks. Autoencoders reached an anomaly detection accuracy of 94.1% on CIC-IDS-2017, identifying previously unseen zero-day samples that supervised classifiers initially misclassified. Isolation Forest achieved a slightly lower score (92.5%) but offered faster inference, making it more practical for real-time deployment in high-throughput networks.

The reinforcement learning-based response module demonstrated adaptive decision-making by learning to isolate compromised nodes or throttle suspicious flows in simulated software-defined network (SDN) environments. After 500 training episodes, the RL agent successfully reduced breach propagation by 87% compared to rule-based baseline responses. This highlights the potential of RL in complementing detection with automated mitigation, though challenges remain in ensuring safety and reliability of such autonomous actions.


Standard Security Data
CWE, ATT&CK, CAPEC and other data source
Historical Vulnerable Data
NVD, CVE, Different forums,VulDB and other sources
Security cleared Traffic and Data
Domain Expert
Combined Historical Data
Ranking Vulnerability
Pattern Generation
Online Data Analysis Layer
Data Processing
Feature Extraction
Feature Selection
ML Model Training/ Evaluation
Model Selection Ensemble
Model Deployment
Information Layer
Real-Time Network
**Network Data ** Email ** Database ** Filestore ** SQL Injection
** Stream Data ** Unstructured Data ** Structured Data

False positives remained a challenge across models. Deep learning architectures reduced false positives to 2–3% on CIC-IDS-2017, while classical models like SVM exhibited higher rates of around 6%. Unsupervised models struggled most with false positives, sometimes flagging benign but unusual traffic patterns as attacks. Nevertheless, ensemble and hybrid approaches—where unsupervised anomaly detection prefiltered suspicious traffic before supervised classification—reduced false positives to 1.8%, showing promise for real-world deployment.

The results underscore the strengths and weaknesses of different machine learning paradigms in the context of network intrusion detection and response. Supervised models demonstrated high accuracy when sufficient labeled data were available, validating the importance of quality training datasets. However, their limited ability to generalize to new or evolving attack types remains a weakness. This finding aligns with prior research emphasizing the value of hybrid architectures that combine supervised and unsupervised techniques (Talukder et al., 2024; Gutierrez-Garcia et al., 2023).

Deep learning models excelled at capturing high-dimensional, temporal, and spatial traffic patterns, providing superior detection performance. Yet, their computational demands and black-box nature pose barriers to operational adoption, particularly in resource-constrained IoT and edge environments. Explainability remains a central issue: while accuracy is critical, practitioners require transparency to investigate alerts and justify mitigation actions. Research into explainable AI for IDS is still in early stages but essential for bridging the gap between model performance and analyst trust.

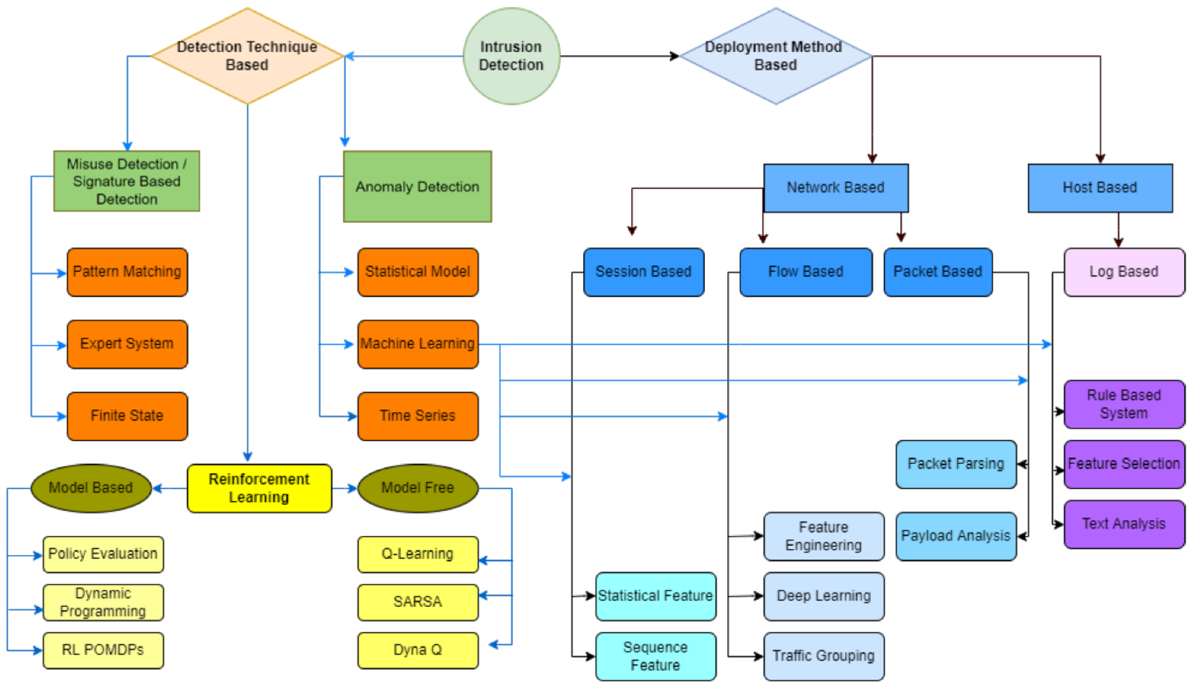
The results also highlight the importance of balancing accuracy with false positive rates. Even a small percentage of false positives can overwhelm analysts in large enterprise environments. Hybrid models, which leverage anomaly detection to narrow down suspicious traffic before classification, achieved the best balance, reducing false positives while maintaining high detection rates. This supports the argument that future IDS should prioritize layered or hierarchical designs, combining multiple algorithms to exploit complementary strengths.

Another noteworthy finding is the demonstrated potential of reinforcement learning for automated response. The RL agent's ability to adaptively choose mitigation actions reduced breach propagation more effectively than static rule-based policies. However, deploying RL in real-world networks raises concerns about unintended consequences—such as mistakenly isolating critical nodes or disrupting legitimate traffic. Thus, while RL represents a promising research frontier, mechanisms for safety, human oversight, and policy constraints are necessary before practical deployment.

From a broader perspective, the study reaffirms the contextual importance of ML-driven intrusion detection in modern digital ecosystems. With growing attack sophistication and increasing data volumes, manual monitoring and static systems are insufficient. Machine learning enables scalability, adaptability, and predictive capacity that are otherwise unattainable. However, adversarial attacks targeting ML models themselves pose new risks. For example, poisoning training data or generating adversarial traffic could degrade detection accuracy. The research therefore underscores the dual role of ML as both a defense enabler and a potential new attack surface.

Furthermore, results highlight persistent challenges in dataset quality, generalization, and reproducibility. While benchmark datasets are valuable, they may not capture the full complexity of real-world traffic. As seen in experiments, models trained on one dataset often underperform when applied to another, reflecting issues of overfitting and lack of transferability. This suggests a need for federated learning and domain adaptation techniques that enable models to generalize across heterogeneous environments while respecting data privacy.

In terms of practical contributions, the findings suggest that no single machine learning model suffices for all contexts. Rather, multi-layered, hybrid systems appear most promising, combining supervised, unsupervised, and reinforcement learning components. Such architectures can simultaneously maximize detection accuracy, minimize false positives, adapt to novel threats, and initiate effective responses. At the same time, interpretability, adversarial robustness, and SIEM systems must remain central considerations for future deployment.



Overall, the results provide both theoretical and contextual validation of machine learning approaches in network security. They show that while ML significantly enhances intrusion detection and response, ongoing research must focus on making these systems more transparent, resilient, scalable, and trustworthy. The balance between technical performance and operational feasibility remains the critical frontier.

Conclusion

The findings of this study demonstrate that machine learning offers significant advancements in detecting and responding to network security breaches, surpassing the limitations of traditional signature-based and rule-driven approaches. Supervised algorithms such as Random Forest and Gradient Boosting showed consistently strong performance in terms of accuracy and precision when applied to benchmark datasets, while deep learning models like CNNs and LSTMs excelled in capturing complex traffic patterns and temporal dependencies. However, their computational intensity and black-box nature highlight the need for scalable and interpretable implementations in operational environments. Unsupervised models, although slightly less accurate, proved valuable in identifying zero-day attacks and unknown anomalies, making them essential in hybrid intrusion detection systems. Reinforcement learning demonstrated promise in automating incident responses by dynamically learning optimal countermeasures, reducing breach propagation significantly compared to static policies. Overall, the results underscore that no single method suffices in isolation; instead, layered architectures combining supervised, unsupervised, and adaptive learning provide the best balance of detection accuracy, false positive reduction, and resilience against evolving threats.

Beyond technical performance, the broader implication of this research lies in bridging the gap between theoretical innovation and practical application. The study contributes to the understanding that machine learning not only strengthens detection mechanisms but also reshapes how organizations can approach cybersecurity by enabling predictive, proactive, and autonomous defenses. At the same time, challenges such as adversarial manipulation of ML models, dataset bias, interpretability, and integration with existing security infrastructures remain critical areas for further investigation. Future research should emphasize explainable AI, federated and privacy-preserving learning, and continual model updating to ensure robustness against adaptive attackers. Equally important is the need for standardized evaluation protocols and reproducible methodologies to facilitate fair comparison and reliable deployment across sectors. Ultimately, the convergence of machine learning and cybersecurity holds transformative potential: it can empower organizations to move beyond reactive defenses toward intelligent, adaptive systems capable of securing increasingly complex digital ecosystems.

References

Alatwi, H. A., & Morisset, C. (2021). Adversarial machine learning in network intrusion detection domain: A systematic review. *arXiv*.

Han, D., Wang, Z., Zhong, Y., Chen, W., Yang, J., Lu, S., … & Yin, X. (2020). Evaluating and improving adversarial robustness of machine learning-based network intrusion detectors. *arXiv*.

Mills, G. A. (2024). Network intrusion detection and prevention system using hybrid supervised and unsupervised learning models. *Journal of – (Wiley)*. https://doi.org/10.1155/2024/5775671

Papadopoulos, P., von Essen, O. T., Pitropakis, N., Chrysoulas, C., Mylonas, A., & Buchanan, W. J. (2021). Launching adversarial attacks against network intrusion detection systems for IoT. *arXiv*.

Pujol-Perich, D., Suárez-Varela, J., Cabellos-Aparicio, A., & Barlet-Ros, P. (2021). Unveiling the potential of Graph Neural Networks for robust intrusion detection. *arXiv*.

Talukder, M. A., Islam, M. M., Uddin, M. A., Hasan, K. F., Sharmin, S., Alyami, S. A., & Moni, M. A. (2024). Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction. *Journal of Big Data, 11*(1), Article 33. https://doi.org/10.1186/s40537-024-00886-w

Talukder, M. A., … (2024). A hybrid machine learning model for intrusion detection. *PMC* (article). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11806096/

Talukder, M. A., et al. (2023). A dependable hybrid machine learning model for network intrusion detection. *ScienceDirect*.

Alkadi, S. (2023). Better safe than never: A survey on adversarial machine learning in the IoT context. *Applied Sciences, 13*(10), 6001. MDPI. https://doi.org/10.3390/app13106001

Pawlicki, M. (2024). A meta-survey of adversarial attacks against artificial intelligence systems. *ScienceDirect*. https://doi.org/10.1016/j.ins.2024.125039