Challenges and Future Directions in Machine Learning for Intrusion Detection Systems (IDS)

Shailendra Kumar, PhD Scholar, Guru Ghasidas Vishwavidyalaya Bilaspur Chhattisgarh
Dr. Shrabanti Mandal, Guru Ghasidas Vishwavidyalaya Bilaspur Chhattisgarh

Abstract

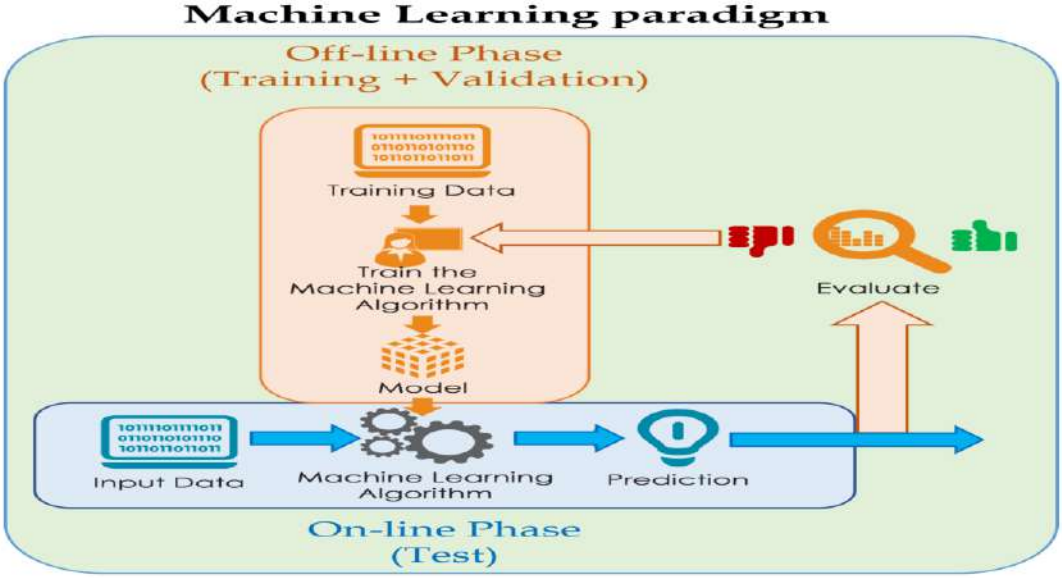
The rise of machine learning (ML) has reshaped many areas of cybersecurity, especially Intrusion Detection Systems (IDS). These systems are essential for spotting suspicious activity and protecting networks from potential attacks. By learning patterns from large volumes of data, machine learning–based IDS can detect unusual behavior and flag possible intrusions, making them more intelligent and adaptive than traditional rule-based systems. Yet, despite their potential, ML-driven IDS still face several practical challenges. Problems such as poor data quality, imbalanced datasets, limited model interpretability, and difficulty adapting to constantly evolving threats can reduce their overall effectiveness. These limitations often make it harder to deploy such systems reliably in real-world environments. This paper takes a close look at these ongoing challenges and reviews current research efforts aimed at overcoming them. It also discusses promising future directions, including hybrid modeling approaches, transfer learning techniques, explainable AI (XAI), and federated learning frameworks. Ultimately, the study highlights the importance of continued innovation in this field to strengthen IDS performance and build more resilient network security systems.

Keywords: Intrusion Detection Systems, Machine Learning, Class Imbalance, Explainable AI, Transfer Learning, Cybersecurity.

1. Introduction

Intrusion Detection Systems (IDS) play a central role in keeping computer networks secure. They are built to spot unauthorized access, suspicious behavior, and unusual patterns in network traffic. However, as cyberattacks have become more sophisticated and unpredictable, traditional IDS approaches—mainly those relying on signature matching or fixed rule sets—have struggled to keep up. These conventional methods work well for known threats, but they often fail when faced with new or evolving attacks.

To overcome these limitations, machine learning (ML) has become an increasingly important part of modern IDS. Unlike rule-based systems, ML-driven models can learn directly from data. They identify patterns, adapt to changing conditions, and improve over time. This allows them to detect not only known threats but also previously unseen or emerging attacks by recognizing unusual behavior in network traffic.



Machine learning–based IDS can analyze complex relationships within large volumes of data, draw insights from past incidents, and classify traffic as either legitimate or malicious. Despite these advantages, integrating ML into IDS is far from straightforward. Several challenges remain. Issues such as poor data quality, imbalanced datasets, and the difficulty of keeping models updated in a constantly evolving threat landscape can significantly affect performance.

$$f(t) = L^{-1}\{F(s)\} = \frac{1}{2\pi i}\lim_{T\to\infty}\int_{\gamma-iT}^{\gamma+iT} e^{st}F(s)ds,$$

$$f(t) = L^{-1}\{F(s)\} = \frac{1}{2\pi i}\lim_{T\to\infty}\int_{\gamma-iT}^{\gamma+iT} e^{st}F(s)ds,$$

$$\lim_{R\to\infty}\int_0^R f(t)e^{-ts}dt$$

$$\int_0^R f(t)e^{-ts} \mid dt$$

$$F(s) = (s - s_0)\int_0^\infty e^{-(s-s0)t}\beta(t)dt,\ \beta(u) = \int_0^u e^{-s0t}f(t)dt.$$

$$\{L^*g\}(s) = \int_0^\infty e^{-st}dg(t).$$

$$g(x) = \int_0^x f(t)dt$$

$$= \int_{-\infty}^{\infty} e^{-\iota\omega t}f(t)dt.$$

In addition, many advanced ML models—especially deep learning systems—operate as "black boxes," making it difficult to understand how they arrive at specific decisions. In security environments, where accountability and trust are essential, this lack of interpretability raises serious concerns. This paper examines the key challenges associated with machine learning–based intrusion detection systems, highlights the limitations of current methods, and outlines possible future directions to make these systems more reliable, transparent, and effective.

2. Previous Work

Signature-based intrusion detection systems (IDS) work by matching incoming network traffic against a database of known attack patterns. They are very effective at catching attacks that have already been identified and documented. However, they struggle with new or zero-day threats because those signatures simply don't exist yet in the database. While these systems are precise when it comes to recognizing known patterns, they depend heavily on regularly updated signature libraries, which can become time-consuming and difficult to maintain.

Anomaly-based IDS take a different approach. Instead of looking for known attack patterns, they first learn what "normal" network behavior looks like and then flag anything that deviates from that baseline. This makes them more flexible and better suited to detecting previously unseen attacks. The downside, however, is that they often generate a higher number of false alarms—especially in complex or constantly changing network environments. Incorporating machine learning, particularly unsupervised techniques, can greatly enhance anomaly detection by uncovering subtle patterns and relationships in the data that traditional methods might miss.

Hybrid IDS attempt to bring the best of both worlds together. By combining signature-based precision with the adaptability of anomaly-based detection, these systems aim to provide stronger and more balanced protection. Machine learning plays an important role here as well, helping hybrid systems analyze data in real time, refine their decision-making processes, and ultimately improve overall detection performance.

3. Motivation

One of the biggest hurdles in applying machine learning to intrusion detection systems (IDS) is the quality of the data used to train the models. In real-world networks, traffic data is rarely clean or perfectly organized. It often contains noise, missing values, or irrelevant information. That's why preprocessing becomes so important. Carefully cleaning the data and selecting meaningful features can make a huge difference. Even the most advanced algorithm won't perform well if it's trained on poorly prepared data.

Another persistent challenge is class imbalance. In most networks, the vast majority of traffic is normal, while malicious activity makes up only a tiny portion. As a result, models tend to favor the majority class and may overlook actual attacks. This leads to false negatives—situations where real threats slip through undetected. Techniques like oversampling, undersampling, and synthetic data generation methods such as SMOTE are commonly used to address this imbalance. However, these solutions are not perfect; they can sometimes introduce overfitting or increase computational complexity.

Interpretability is also a major concern. Many machine learning models, particularly deep learning systems, function as "black boxes." They produce predictions, but it's not always clear how they arrived at those decisions. In the context of IDS, this lack of transparency can be problematic. Security analysts need to understand why certain traffic was flagged as malicious in order to trust and act on the system's alerts. This has led to growing interest in explainable AI (XAI), which focuses on developing models that can provide clear, human-understandable explanations for their decisions.

Adding to these challenges is the constantly evolving nature of cyber threats. Attack techniques change rapidly, and models trained on historical data may struggle to recognize new or previously unseen threats. The decentralized and dynamic landscape of cybersecurity makes it difficult to build models that remain effective over time. Approaches like transfer learning—where knowledge from one domain is applied to another—and online learning, which allows models to update continuously as new data becomes available, offer promising ways to improve adaptability.

Finally, computational demands cannot be ignored. Many advanced machine learning approaches, especially deep learning models, require substantial processing power and memory for training and deployment. This creates difficulties when implementing IDS in resource-constrained environments or systems that require real-time detection. Balancing model accuracy with efficiency remains an ongoing challenge in the development of machine learning–based intrusion detection systems.

4 Data Acquisition

Supervised learning techniques—such as Support Vector Machines (SVMs), Random Forests, and Decision Trees—are commonly used in intrusion detection systems (IDS) to sort network traffic into either malicious or legitimate categories. These models rely on labeled training data, which can be a drawback in fast-changing environments where new types of attacks appear regularly. Even so, when the dataset is well-prepared and properly balanced, supervised methods tend to deliver strong and reliable accuracy.

$$\lim_{\sigma\to 0^+} F(\sigma+i\omega)=\hat{f}(\omega)$$

$$G(s)=M\{g(\theta)\}=\int_0^\infty \theta^s g(\theta)\frac{d\theta}{\theta}$$

$$\Delta_T(t)\stackrel{def}{=}\sum_{n=0}^{\infty}\delta(t-nT)$$

$$x_q(t) \overset{def}{=} (t)\Delta_T(t) = x(t)\sum_{n=0}^{\infty}\delta(t-nT)$$

$$= \sum_{n=0}^{\infty} x(nT)\delta(t-nT) = \sum_{n=0}^{\infty} x[n]\delta(t-nT)$$

$$X_q(s) = \int_{0^-}^{\infty} x_q(t)e^{-st}dt$$

$$= \int_{0^-}^{\infty}\sum_{n=0}^{\infty} x[n]\delta(t-nT)e^{-st}$$

$$= \sum_{n=0}^{\infty} x[n]\int_{0^-}^{\infty}\delta(t-nT)e^{-st}$$

$$= \sum_{n=0}^{\infty} x[n]e^{-nsT}$$

$$X(z) = \sum_{n=0}^{\infty} x[n]z^{-n}$$

Unsupervised learning approaches, including clustering methods like K-means and anomaly detection models such as Isolation Forests, are especially helpful when labeled data is limited or unavailable. Instead of depending on predefined categories, these techniques learn patterns directly from the data. This makes them particularly useful for spotting new or previously unseen attacks. The main challenge, however, is keeping false alarms under control, since the model may sometimes misclassify normal traffic as suspicious.

Deep learning methods, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have also gained attention for IDS applications. They are well suited to handling large volumes of network traffic and can capture complex patterns and relationships within the data. Their ability to learn layered feature representations makes them powerful tools for detecting sophisticated threats. That said, they come with higher computational costs and typically require large amounts of labeled data to perform effectively.

Reinforcement learning (RL) has emerged as another promising direction, particularly for adaptive IDS. In this approach, the system learns by interacting with its environment and adjusting its actions based on feedback from network activity. Over time, an RL-based IDS can adapt to evolving threats by continuously learning from new data, offering a flexible and dynamic response to emerging security challenges.

5. Methodology

Researchers are increasingly turning to hybrid approaches to make intrusion detection systems (IDS) more accurate and adaptable. Instead of relying on a single model, these approaches combine different machine learning techniques to play to their strengths. For instance, supervised learning can be used to identify known attack patterns, while unsupervised learning helps uncover new or previously unseen threats. Together, this combination can lower false alarms and improve the system's ability to catch emerging attacks.

Transfer learning also offers promising possibilities. By taking a model that has already been trained on one dataset or task and adapting it to a new one with only limited additional data, IDS can respond more quickly to newly emerging threats. In environments where attack methods constantly evolve, this flexibility is especially valuable. Similarly, online learning allows models to update themselves continuously as new data flows in. This means the IDS can adjust in real time, staying aligned with changing network behavior rather than becoming

outdated.

Another important direction involves integrating external threat intelligence and contextual information into IDS models. When systems draw on data such as known vulnerabilities, established attack patterns, and real-time network activity, their detection capabilities become much stronger. This added context helps the IDS recognize more sophisticated and subtle attacks that might otherwise slip through unnoticed.

6 Result

As the number of IoT devices continues to rise, the need for decentralized intrusion detection systems (IDS) has become more pressing. Instead of relying on a central server, these systems can operate directly at the network's edge. Edge computing makes it possible to process data in real time, right where it is generated. This reduces delays, improves response times, and is especially valuable in settings with limited bandwidth or in applications where immediate action is critical.

In security contexts, trust matters. That's why explainable AI plays such an important role. It's not enough for a model to flag a threat—it also needs to explain why it made that decision. Researchers are actively working on developing machine learning models for IDS that are both accurate and interpretable, so security teams can understand, verify, and confidently act on the system's predictions.

$$M[x^n]=\mu_n, \qquad \sum_{m=0}^{k} M[\lambda_{k,m}(x)]=\sum_{m=0}^{k}\lambda_{k,m}=\mu_0$$

$$M[P_n(x)]=\int_0^1 P_n(t)\,d\alpha(t).$$

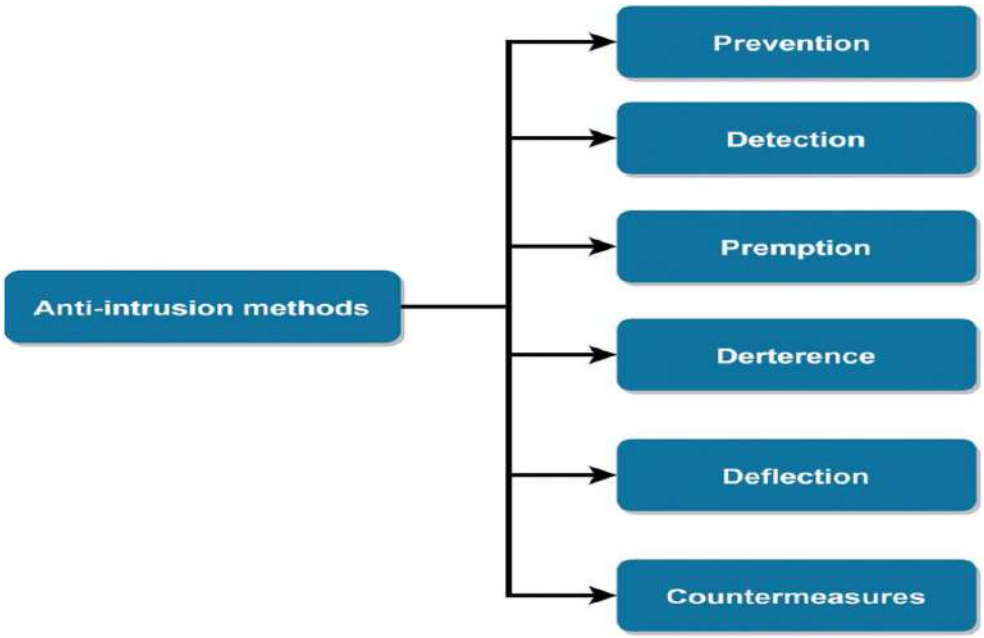
$$\mu_n=M[x^n]=\sum_{m=n}^{k}\frac{m(m-1)\ldots(m-n+1)}{k(k-1)\ldots(k-n+1)}\lambda_{k,m}$$

$$=\sum_{m=n}^{k}\left\{\frac{ky(ky-1)\ldots(ky-n+1)}{k(k-1)\ldots(k-n+1)}-y^n\right\}\lambda_{k,m}$$

$$\mu_n=\lim_{k\to\infty}\int_0^1 t^n\,d\alpha_k(t)$$

$$=\lim_{i\to\infty} n\int_0^1 t^{n-1}[\alpha_{ki}(1)-\alpha_{ki}(t)]dt$$

The integration of machine learning into Intrusion Detection Systems marks a decisive shift in cybersecurity philosophy from static, rule-driven protection mechanisms toward adaptive, intelligence-driven defense architectures. Traditional signature-based IDS were fundamentally reactive in nature, relying on previously observed attack patterns and handcrafted rules. Although effective against known threats, these systems failed to generalize when confronted with zero-day exploits, polymorphic malware, and rapidly evolving attack vectors. Machine learning introduces the capability to learn statistical regularities, infer hidden structures within network traffic, and autonomously adapt to new behaviors. Consequently, ML-based IDS offer enhanced pattern recognition, behavioral modeling, and predictive detection capacities that extend far beyond the operational scope of classical techniques. However, despite these theoretical and empirical advantages, the translation of machine learning from laboratory prototypes to mission-critical production environments remains fraught with complexities. The discussion of these complexities reveals that technological innovation alone is insufficient; practical deployment demands robustness, scalability, trustworthiness, and resilience against intelligent adversaries.



One of the most persistent and foundational obstacles concerns the imbalance and quality of IDS datasets. Real-world network environments exhibit an overwhelming predominance of legitimate traffic, while malicious activities constitute only a small fraction of total events. From a machine learning perspective, this imbalance introduces systematic bias during training, as models tend to optimize for majority classes. Classifiers may therefore achieve deceptively high accuracy by simply predicting benign behavior while failing to detect rare but critical intrusions. This phenomenon is particularly dangerous in cybersecurity because the cost of false negatives is significantly higher than the cost of false positives. Although several mitigation strategies have been proposed, including oversampling, undersampling, and synthetic data generation methods such as SMOTE, these solutions are not without drawbacks. Artificially generated samples may distort the natural distribution of attack behaviors, resulting in models that perform well on curated datasets but poorly in real deployments. Moreover, intrusion data frequently contains noise, incomplete labeling, and outdated attack scenarios, which further degrade model generalization. The discussion therefore underscores that improving IDS effectiveness requires not only algorithmic advancements but also the creation of high-quality, continuously updated, and realistically representative datasets. Without reliable data foundations, even the most sophisticated learning architectures will produce unreliable outcomes.

Closely connected to data-related concerns is the issue of adversarial vulnerability. Unlike many application domains of machine learning, cybersecurity involves an active adversary who deliberately attempts to deceive the detection system. Attackers may analyze IDS decision boundaries and craft adversarial inputs that exploit model weaknesses. Even minimal perturbations in packet headers, timing characteristics, or payload features can lead to misclassification, allowing malicious traffic to bypass defenses. This adversarial dynamic challenges the assumption that learned patterns remain stable over time. Instead, intrusion detection becomes an ongoing strategic contest between defensive learning mechanisms and offensive evasion techniques. The fragility of many deep learning models to adversarial manipulation raises serious concerns regarding their reliability in high-stakes environments such as critical infrastructure, financial networks, or national security systems. Addressing this vulnerability demands a paradigm shift toward robust learning. Techniques such as adversarial training, which exposes models to maliciously perturbed samples during training, and secure

feature engineering, which emphasizes invariant behavioral characteristics rather than easily manipulated surface attributes, represent promising directions. Additionally, research into defensive distillation and certified robustness aims to provide mathematical guarantees of resistance against certain classes of attacks. Nevertheless, achieving comprehensive resilience remains an open challenge, requiring interdisciplinary collaboration between machine learning researchers and cybersecurity experts.

Scalability and real-time processing requirements introduce another layer of complexity that often receives insufficient attention in academic studies. Modern networks generate terabytes of traffic daily, and enterprise or cloud infrastructures may handle millions of events per second. Machine learning models must therefore operate under strict latency constraints, processing streaming data without introducing unacceptable delays. Many state-of-the-art deep learning architectures, while highly accurate, involve substantial computational overhead and memory consumption. Training such models may require specialized hardware accelerators, and inference may still be too slow for real-time detection scenarios. This mismatch between algorithmic sophistication and operational feasibility presents a critical bottleneck for practical deployment. Emerging solutions seek to distribute computation across edge devices, leverage parallel processing frameworks, or employ lightweight models that trade minimal accuracy for significant efficiency gains. Edge computing, in particular, enables preliminary analysis closer to data sources, reducing transmission latency and bandwidth consumption. However, distributing detection responsibilities across multiple nodes introduces challenges related to synchronization, consistency, and security of model updates. Thus, scalable IDS architectures must balance competing demands of speed, accuracy, and resource efficiency, highlighting the necessity of holistic system design rather than isolated algorithmic improvements.

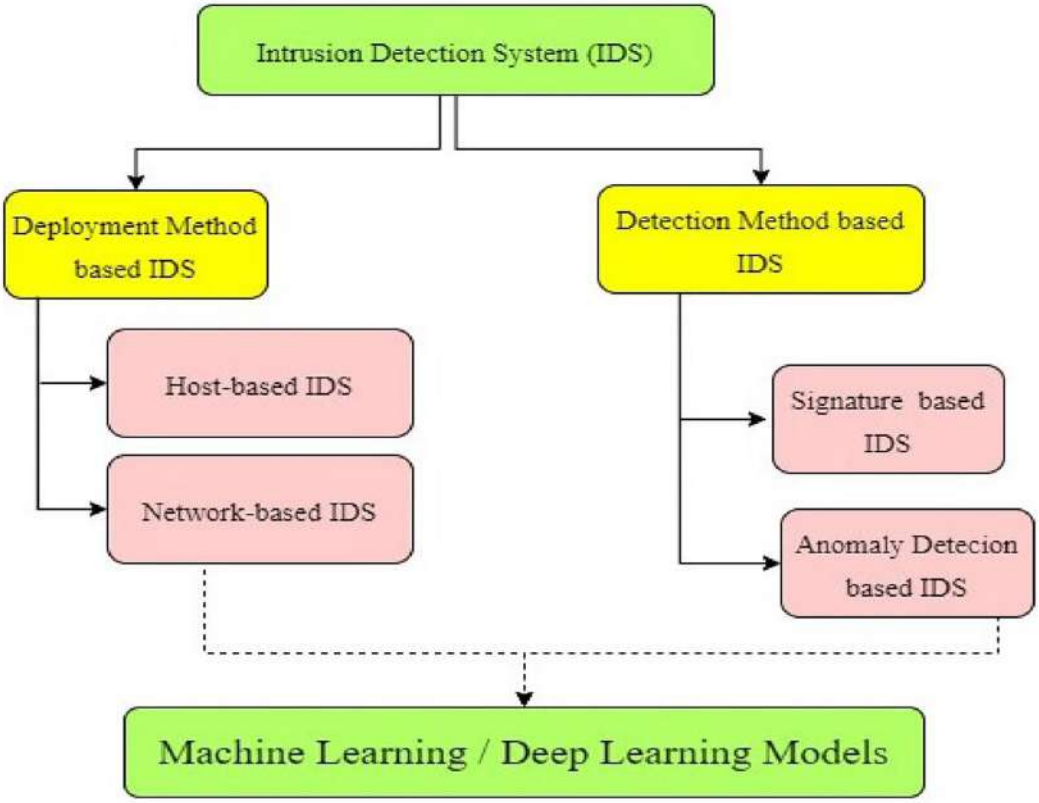
Beyond these immediate technical challenges, broader considerations of interpretability and trust also shape the future trajectory of ML-based IDS. Security analysts and system administrators require clear explanations for detection decisions to assess risk and implement countermeasures. Many advanced models, especially deep neural networks, function as opaque black boxes, providing predictions without transparent reasoning. This opacity undermines trust, complicates debugging, and may hinder compliance with regulatory standards. Explainable Artificial Intelligence has therefore emerged as a critical research direction. By revealing which features or behaviors contributed to an alert, XAI techniques enable human experts to validate decisions and refine defensive strategies. Interpretability is particularly important in cybersecurity, where false alarms can disrupt operations and excessive automation without accountability may introduce unforeseen vulnerabilities. Enhancing transparency ensures that machine intelligence complements rather than replaces human judgment.

Privacy considerations further motivate the exploration of federated learning paradigms. In many settings, particularly healthcare, finance, or government sectors, sharing raw network data across organizations is restricted due to legal and ethical constraints. Federated learning allows multiple entities to collaboratively train shared models without exposing sensitive information, thereby preserving privacy while benefiting from collective knowledge. This decentralized approach also improves generalization by incorporating diverse traffic patterns from multiple environments. Nevertheless, federated systems introduce new challenges such as communication overhead, heterogeneity of local data distributions, and risks of poisoning attacks. Consequently, future IDS research must carefully design secure aggregation protocols and robust update mechanisms to ensure reliability.

Finally, the dynamic nature of cyber threats necessitates continuous learning capabilities. Static models trained once and deployed indefinitely quickly become obsolete as attackers adapt their tactics. Continuous or online learning frameworks enable IDS to evolve in response to new data streams, maintaining relevance over time. However, such adaptability must be balanced

against the risk of catastrophic forgetting, where models lose knowledge of previously learned patterns. Developing mechanisms that preserve historical knowledge while integrating new information remains a crucial research problem. Lifelong learning and incremental updating strategies may provide viable solutions.

Taken together, the discussion reveals that the future of machine learning-based intrusion detection lies not in isolated algorithmic breakthroughs but in integrated, resilient, and adaptive ecosystems. Effective IDS must combine robust data engineering, adversarial defenses, scalable architectures, explainable reasoning, privacy preservation, and continuous learning into a coherent framework. Only through such a multidimensional approach can machine learning realize its transformative potential and deliver reliable protection against the increasingly sophisticated threats of the digital era. Federated learning offers another promising direction. It enables multiple devices or organizations to jointly train a shared machine learning model without exchanging sensitive data. For IDS, this is particularly beneficial. Organizations can improve detection capabilities through collective learning while still protecting the privacy of their network traffic.



Finally, combining machine learning-based IDS with automated incident response systems can dramatically shorten the gap between detecting a threat and responding to it. Automated orchestration allows systems to act quickly and consistently, helping organizations contain and mitigate attacks in real time rather than reacting after the damage is done.

7. Conclusion

Machine learning has completely transformed the way intrusion detection systems work. It has opened the door to smarter, more adaptive methods that can spot not just known threats, but also new and increasingly sophisticated attacks. That said, the journey hasn't been without obstacles. Issues like poor data quality, imbalanced datasets, and the difficulty of interpreting complex models still pose real challenges. Looking ahead, researchers need to focus on developing hybrid approaches that combine the strengths of different models, along with transfer learning techniques that help systems adapt more quickly to new environments. Just

as importantly, explainable AI must be prioritized so that security professionals can understand and trust the decisions these systems make. Bringing in contextual data, leveraging edge computing, and adopting federated learning frameworks also offer promising ways to strengthen intrusion detection in dynamic, distributed network settings. As cyber threats continue to grow in scale and complexity, machine learning–based intrusion detection systems will become even more essential in safeguarding modern networks.

7. References

1. Ni, M. "A review on machine learning methods for intrusion detection system." Applied and Computational Engineering, vol. 27, 2023 — A comprehensive survey of ML methods for IDS.
2. Applied and Computational Engineering
3. Raj, S., Jain, M., & Kamble, M. "A review on intrusion detection system based on various learning techniques." Indian Journal of AI and Neural Networking, 2023 — Examines deep and machine learning for IDS.
4. Khan, N. W., Alshehri, M. S., Khan, M. A., & Almakdi, S. "A hybrid deep learning-based intrusion detection system for IoT networks." Mathematical Biosciences and Engineering, 2023 — Proposes DL-ML models for IoT intrusion detection.
5. Mohammed, M. S., & Talib, H. A. "Using machine learning algorithms in intrusion detection systems: a review." Tikrit Journal of Pure Science, 2024 — Focuses on challenges like feature selection and resource constraints.
6. Mareedu, A. "Machine learning applications in intrusion detection: a comprehensive review." International Journal of Multidisciplinary on Science and Management, 2024 — Broad review of ML techniques in IDS.
7. Review of machine learning algorithm for intrusion detection system, ComniTech: Journal of Computational Intelligence and Informatics, 2024 — Survey of ML algorithms for IDS
8. Rehman, H. M. R. U., et al. "A systematic literature study of machine learning techniques based intrusion detection: datasets, models, challenges, and future directions." Journal of Big Data (2025) — Discusses IDS gaps and future research.
9. Advancements in machine learning-based intrusion detection in IoT: research trends and challenges. Algorithms (MDPI) — Reviews ML/DL advances particularly for DDoS and IoT IDS.
10. "Unveiling machine learning strategies and considerations in intrusion detection systems: a comprehensive survey." Frontiers in Computer Science — Highlights DL vs ML trade-offs and dataset limitations.
11. Pinto, A., Herrera, L.-C., Donoso, Y., & Gutierrez, J. A. "Survey on intrusion detection systems based on machine learning techniques for the protection of critical infrastructure," Sensors, 2023 — Focus on ML-IDS for infrastructure.
12. Chua, T.-H., & Salam, I. "Evaluation of machine learning algorithms in network-based intrusion detection system." arXiv preprint, 2022 — Empirical evaluation of ML models for IDS tasks.
13. Nakıp, M., & Gelenbe, E. "Online self-supervised deep learning for intrusion detection systems." arXiv preprint, 2023 — Self-supervised DL for IDS.
14. Yuan, X., Han, S., Huang, W., Ye, H., Kong, X., & Zhang, F. "A simple framework to enhance the adversarial robustness of deep learning-based intrusion detection system." arXiv preprint, 2023 — Proposed robust DL-ML IDS hybrid.
15. "Machine learning based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction." Journal of Big Data (2024) — Novel methods to address imbalance and dimensionality in IDS data.

16. Yakub Reddy, K., & ShankarLingam, G. "Artificial intelligence in intrusion detection systems: trends, frameworks, and future directions for cybersecurity." International Journal of Intelligent Systems and Applications in Engineering, 2024 — AI-IDS frameworks and trends.
17. Ahmed, U. et al. "Signature-based intrusion detection using machine learning and deep learning approaches empowered with fuzzy clustering." Scientific Reports, 2025 — Ensemble and fuzzy cluster methods in IDS. (cited in survey)
18. Kikissagbe, B. R., & Adda, M. "Machine learning-based intrusion detection methods in IoT systems: a comprehensive review." Electronics (MDPI), 2024 — IoT IDS ML review focusing on practical challenges.
19. Chowdhury, Z. A., Rahman, M. M., & Azhar, T. "Advances in intrusion detection systems: integrating machine learning, deep learning, IoT, and federated learning." International Journal of Computer Applications, 2024 — Addresses integration of ML and federated learning for distributed IDS.
20. Hozouri, A. "A comprehensive survey on intrusion detection systems with machine learning and deep learning." Springer Nature (2025) — Consolidates architectural frameworks, models, and datasets for ML/DL IDS research.
21. A review of various datasets for machine learning intrusion detection systems, International Journal of Intelligent Systems and Applications in Engineering, 2024 — Examines dataset challenges for ML-IDS research.