The Bioscan AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES 20(4): 1775-1791, 2025 www.thebioscan.com

Machine Learning-Driven Data Mining Techniques for Enhancing Cyber Security

Ms. M. Sri Soundharyaa[1], Dr. D. Kavitha[2]
[1]Research Scholar, Department of Computer Science, Nandha Arts and Science College (Autonomous), Erode – 52, TN, India Email: srisoundharyaaprabhu@gmail.com
[2]Assistant Professor and Head, Department of Information Technology, Nandha Arts and Science College (Autonomous), Erode – 52, TN, India
Email: kavitha.d@nandhaarts.org

DOI: 10.63001/tbs.2025.v20.i04.pp1775-1791

Keywords
Cybersecurity, Data Mining, Machine Learning, Intrusion Detection, Anomaly Detection, Supervised Learning, Unsupervised Learning, Cyber threat, Adversarial Attack.

Received on:
25-10-2025

Accepted on:
17-11-2025

Published on:
30-12-2025

ABSTRACT

As cyberattacks become more complex and frequent, cybersecurity has emerged as a top research priority. Traditional signature-based defense systems struggle to identify new and evolving threats. To tackle this challenge, researchers are increasingly turning to machine learning and data mining techniques to uncover patterns, detect suspicious activities, and reveal hidden relationships within vast security datasets. This paper will take a deep dive into various data mining methods, including classification, clustering, association rule mining, and anomaly detection, and how they can be applied to enhance cybersecurity. Additionally, it will explore the integration of supervised, unsupervised, and ensemble learning approaches, such as Support Vector Machines, random forests, neural networks, and deep learning architectures, particularly in the contexts of intrusion detection, malware analysis, and fraud detection. The discussion will also cover benchmark datasets like KDD Cup 99, NSL-KDD, CICIDS 2017, and UNSW-NB15, along with performance metrics such as accuracy, precision, recall, F1-score, and AUC. Furthermore, the paper will address emerging challenges in the field, including class imbalance, concept drift, adversarial attacks, and scalability. Finally, it will outline promising research directions that focus on hybrid intelligent systems, explainable AI, and big data-driven cybersecurity analytics.

Introduction

In our increasingly interconnected digital world, cyber threats are becoming more intricate, frequent, and damaging. The traditional rule-based security measures are falling short against sophisticated attacks like zero-day exploits, advanced persistent threats (APTs), and polymorphic malware. This has led to a growing need for smart and adaptable security systems. Enter data mining and machine learning (ML), which have proven to be invaluable in the realm of cyber security. These technologies can reveal hidden patterns, spot anomalies, and forecast potential threats within vast and complex datasets. Machine learning-driven data mining techniques provide a forward-thinking approach to cyber defense by learning from past data and pinpointing deviations from typical behavior. Techniques such as supervised, unsupervised, and reinforcement learning can be utilized for various tasks, including intrusion detection, malware classification, phishing detection, and fraud prevention. These models are constantly evolving and improving as they analyze real-time data streams, making them exceptionally effective at identifying new attack vectors that traditional systems might overlook.

Additionally, combining ML with big data analytics allows security systems to scale efficiently in ever-changing environments like cloud computing and the Internet of Things (IoT). However, while the potential of machine learning in cyber security is significant, it also comes with its own set of challenges, such as data quality, model interpretability, and the threat of adversarial attacks. Nevertheless, ongoing progress in feature engineering, ensemble learning, and deep learning is helping to address these hurdles.By using advanced machine learning and data mining techniques, organizations can greatly improve their cyber security. This means moving away from just reacting to threats and instead adopting smart, data-driven strategies that can predict and neutralize risks before they lead to any harm.

Cybersecurity and Datamining

Cybersecurity has become a critical pillar of modern digital ecosystems due to the exponential growth of data, the ubiquity of interconnected devices, and the sophistication of cyberattacks. With the proliferation of cloud services, Internet of Things (IoT) networks, and mobile computing, organizations are increasingly vulnerable to security breaches that can compromise confidentiality, integrity, and availability of information. Reports from international agencies consistently indicate a steep rise in cybercrime, including ransomware, phishing, distributed denial-of-service (DDoS) attacks, and advanced persistent threats (APTs). These attacks not only inflict substantial financial losses but also erode trust in digital infrastructure and compromise sensitive personal and organizational data.[13]

The importance of cybersecurity extends beyond protecting individual organizations. At the societal level, cyber defense is integral to safeguarding national security, critical infrastructures such as power grids, healthcare systems, and financial institutions. A single breach in these domains can disrupt essential services, trigger cascading failures, and cause irreversible socio-economic consequences. As digital transformation accelerates, ensuring robust and adaptive cybersecurity strategies becomes indispensable to maintaining digital resilience and sustainable technological progress.

Traditional cybersecurity models primarily rely on signature-based detection and predefined rules. Although effective against known threats, these methods are insufficient in combating zero-day exploits and polymorphic attacks that continuously mutate their signatures. Consequently, there is a pressing need for intelligent systems capable of learning from large volumes of dynamic data, recognizing hidden attack patterns, and adapting to emerging threat landscapes. This is where data mining and machine learning provide significant advancements.

Data mining and machine learning (ML) have emerged as transformative approaches in cybersecurity, owing to their ability to analyze massive, high-dimensional, and heterogeneous security datasets. Unlike traditional approaches, data mining focuses on extracting implicit patterns, correlations, and anomalies from structured and unstructured data sources such as network traffic logs, system events, and user behavior profiles. By discovering relationships that are not explicitly encoded, data mining techniques provide

the foundation for proactive threat identification and mitigation.[11]

Classification methods such as Decision Trees, Support Vector Machines (SVM), and Random Forests enable the identification of malicious versus benign activities with high accuracy. Clustering techniques like k-means and DBSCAN facilitate the grouping of anomalous behavior without prior labeling, making them particularly useful for detecting novel attacks. Association rule mining algorithms such as Apriori and FP-growth reveal frequent co-occurrence patterns between features, which can uncover attack signatures embedded within large datasets. Furthermore, anomaly detection techniques, including density-based models and autoencoders, assist in identifying deviations from normal network behavior, which is crucial for recognizing stealthy intrusions.

Machine learning enhances these data mining methods by providing adaptive models that improve detection accuracy over time. Supervised learning algorithms leverage labeled datasets to build predictive models, while unsupervised learning algorithms are instrumental in handling unlabeled or partially labeled cybersecurity data. Semi-supervised and ensemble learning methods further bridge gaps by improving robustness against noisy and imbalanced datasets.

Recent advancements in deep learning architectures—such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Transformer-based models—have enabled sophisticated feature extraction and temporal-spatial correlation analysis in intrusion detection and malware classification. For example, LSTM networks effectively capture sequential dependencies in network traffic, thereby identifying complex multi-stage attacks. Moreover, graph-based learning approaches and federated learning paradigms are emerging as powerful tools to detect distributed attacks in decentralized environments.

By integrating data mining and ML, cybersecurity systems evolve from reactive to proactive defense strategies. They enable real-time monitoring, automated detection of previously unseen attack vectors, and scalable analysis of large security datasets. This paradigm shift is essential in addressing the dynamic and adversarial nature of cyber threats.

This study's impetus comes from the urgent need to develop intelligent cybersecurity frameworks that can cope with the growing complexity, scale, and sophistication of modern cyberattacks. Existing systems, although valuable, often suffer from high false alarm rates, inability to generalize across diverse attack scenarios, and limited scalability in big data environments. Additionally, the increasing frequency of zero-day exploits necessitates detection models that go beyond static rule sets and embrace adaptive learning mechanisms.

This research aims to bridge these gaps by systematically exploring the integration of data mining and machine learning techniques within cybersecurity applications. This study makes three contributions:

Comprehensive Analysis of Data Mining Techniques: The paper reviews the

application of classification, clustering, association rule mining, and anomaly detection methods in intrusion detection, malware analysis, phishing detection, and fraud detection. This analysis highlights the strengths and limitations of each technique in addressing different attack scenarios.

1. **Evaluation of Machine Learning Models in Cyber Defense:** The study investigates the role of supervised, unsupervised, and deep learning methods in enhancing cybersecurity resilience. Emphasis is placed on how hybrid and ensemble approaches can mitigate challenges such as class imbalance, high-dimensionality, and adversarial evasion.
2. **Identification of Open Challenges and Research Directions:** By examining current limitations—including concept drift, scalability for big data, interpretability, and adversarial robustness—the paper outlines promising research avenues. These include the adoption of explainable AI, integration with big data analytics platforms, and the development of real-time, hybridized detection systems.

Overall, this research contributes to advancing the field by synthesizing existing methodologies, identifying unresolved challenges, and proposing pathways for the next generation of data-driven, intelligent cybersecurity solutions. By leveraging the synergies between data mining and machine learning, the study aims to strengthen defenses against evolving threats and foster secure, resilient digital ecosystems.

Intrusion Detection System

Intrusion is anything that compromises the confidentiality, availability, and integrity of the information. Usually, the cloud doesn't allow the usage of vulnerability scanners to detect the intrusion as it distinguishes between the normal and abnormal traffic complex. As the name indicates, the Intrusion Detection System (IDS) is an application that mainly identifies the intrusions in the cloud network and protects it from an intruder at an early stage. The distributed nature of the cloud makes it an easier target for intruders to exploit its weakness. The impact that intrusions make in the cloud environment is usually very high. This increases the challenges associated with securing the data and system in the cloud [12].

The main aim of the IDS is to identify the intrusions in the early stage and prevent the system from severe consequences. With the IDS software's help, one can identify the attack at an initial stage and notify the security personnel immediately about its occurrence. The IDS helps to prevent the cloud environment from various attacks and threats. The main disadvantage associated with the IDS system is the high false alarm rate associated with it. Because if a false alarm is received, the security personnel terminate every communication immediately concerning it as an attack attempt that results in the cloud server's downtime.

The users often doubt the cloud security due to the lack of complete control over their data; in this case, the IDS can be of great use in protecting the users' information assets. The attacks that take place in the cloud is of two forms: insider and outsider attacks. Attacks that originate in the external environment are called external

attacks. The attacks that occur within the organization are called insider attacks where the unauthorized user tries to gain access to the cloud user devices. In this scenario, the IDS monitors the cloud network to identify malicious activity, entry, and illegal file modifications. Incidents are the attacks that occur in distinctive groups.

Most of the incidents that take place in the cloud are malicious whereas the others or not. For instance, a user who accidentally types another address in place of their address of the computer makes illegal access to another system without authorization. The four main essential functionalities of the IDS system are surveillance, identification, investigation, and acknowledgment of the intrusion events. The malicious intent of the attacker is mainly identified by analyzing the acquired data. Network-based, host-based, hypervisor-based, and application-based are the four types of IDS monitoring environments [13].

IDS Type	**Monitoring Environment**	**Key Characteristics**	**Deployment Example**
Network-based IDS	Monitors network traffic across subnets or segments	Analyzes all network packets in real-time, detects suspicious traffic, uses signature/anomaly-based detection. Positioned at network chokepoints like firewalls or routers.	Sensors placed in subnet with firewall, monitors network packets
Host-based IDS	Installed on individual hosts (servers, endpoints)	Monitors system activities, file integrity, system logs analysis on the host itself. Detects unauthorized changes or behavior at the device level.	Agent on a critical server or endpoint to analyze local system activities
Hypervisor-based IDS	Monitors virtualized environments at the hypervisor level	Observes traffic and activities between virtual machines (VMs) and the hypervisor layer. Provides visibility into VM-to-VM	Deployed on virtualization hosts to monitor VMs and hypervisor interactions [general knowledge, typical feature]

IDS Type	Monitoring Environment	Key Characteristics	Deployment Example
		and VM-to-network activity.	
Application-based IDS	Monitors specific applications or application protocols	Examines application-level data and behaviors such as SQL queries, APIs, or web application traffic. Detects attacks targeted at applications (e.g., SQL injection).	Deployed inside web servers or application middleware to track app-specific protocols

Table 1: IDS types and characteristic

This table provides a clear comparison of the four IDS monitoring environments by their scope, typical deployment, and monitoring focus.

- Network-based Intrusion Detection System: In this situation, network traffic is monitored for particular network components or devices, and then the traffic is analyzed for malicious behaviour. To identify different types of intrusions, both the internal and external traffic towards the network is analyzed. After the attack is detected, it takes every necessary step to safeguard the cloud from attacks. The cloud provider is the one who controls and implements the network-based IDS in the cloud. They are also capable of identifying the attacks that take place in the virtual machine and hypervisor and not capable of identifying the attack that occurs inside the virtual network.
- Host-based Intrusion Detection System: The host's state and dynamic behaviour (computer) is monitored here. It mainly identifies the resources that the host application use. Some hybrid approaches integrate both the network and host-based IDS to form a hybrid IDSD which offers increased flexibility. By monitoring the inbound and outbound packets, the malicious activity is identified and an alert is given to the cloud provider. The information obtained contains the essential details such as which attack pattern and which users were targeted during this attack. The host-based IDS can be implemented in both the virtual machine and host machine. The users can also control the host-based IDS that are deployed either in the host or virtual machine. The cloud providers are the ones who can monitor the host-based IDS that is deployed in the hypervisor.
- Hypervisor-based Intrusion Detection System: The hypervisor

is also known as a virtual machine monitor that creates and runs the virtual machines. The host machine is the one in which the hypervisor runs one or more virtual machines and the virtual machine is known as the guest machine. The hypervisor-based IDS is implemented between the abstract and its underlying host to mainly add security to the cloud. In this way, the kernel is free of any threats and the hypervisor-based IDS analyses the system metrics from cloud components to identify any negative patterns.

Intrusion Detection Systems (IDS) are vital components in cloud and network security, aimed at identifying unauthorized access or malicious activities. Among them, Application-based IDS focuses on monitoring specific applications by analyzing their log files. It gathers input directly from application data sources, offering targeted detection but is limited to the particular application being monitored. In real-time intrusion detection, systems continuously monitor network or system activity to detect and alert users about ongoing intrusions as they occur. This allows for immediate response and mitigation. Real-time IDS can also function in offline mode, analyzing historical or audit data to uncover past intrusions, although this method introduces a delay due to the non-instantaneous nature of analysis. The audit data may originate from distributed or centralized sources and involves post-event processing.

Intrusion detection methodologies are categorized into three main models:

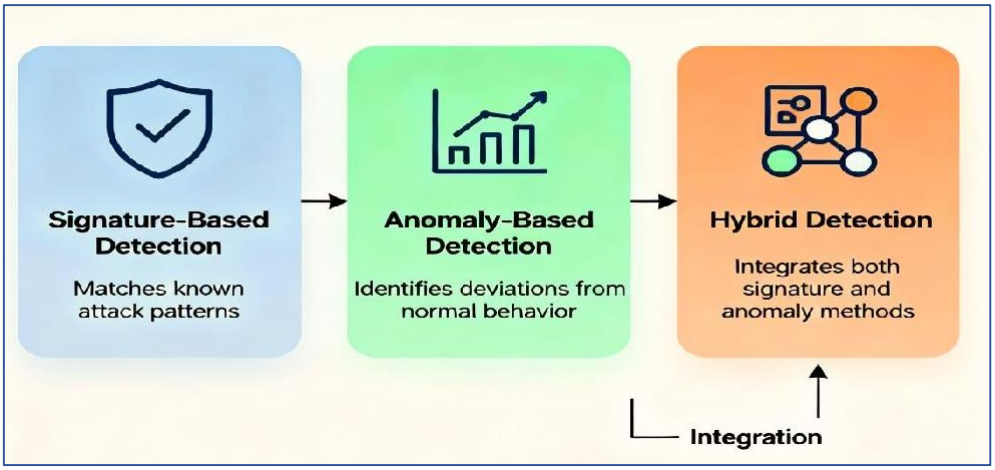

Fig 1: Intrusion detection methodologies

- **Signature- Based Detection:** This model is based on predefined signatures or patterns of known attacks. It compares current activity and a database of attack signatures. If a match is found, the system denotes it as an intrusion. While effective against known threats, it won't detect novel or previously unknown attacks.
- **Anomaly-Based Detection:** This approach defines a baseline for normal behaviour and identifies deviations from it as potential intrusions. It can detect unknown attacks but may suffer from higher

false-positive rates due to variability in legitimate user behaviour.

- **Hybrid Detection:** Combining misuse and anomaly detection, the hybrid model contains the strengths of both techniques. It is designed to detect both known and unknown intrusions by employing signature-based recognition for familiar attacks and behaviour-based analysis for new or evolving threats.

Overall, integrating real-time monitoring with hybrid detection strategies enhances the effectiveness of IDS in identifying a broader range of cyber threats in cloud environments [15].

Analysis of Cybersecurity and Datamining

The intrusion detection process with possible intrusions detections is automated by IDS software. The most critical intrusion detection tool proves to prevent network infrastructure and detect network intrusions from internal IDS attacks, cybercrimes, and web threats. A most effective security measure was proposed by IDS [15]. The process of monitoring networks or computers is called Intrusion detection, and an intrusion is an abnormal activity or file modification caused by an unauthorized entry. Therefore, an efficient IDS is free from false errors with an overhead as low as possible, available without interruption, challenging to cheat, easily configurable, fault-tolerant, self-monitored, and fast. It aims to evaluate information systems and to perform early detection of malicious activity for reducing the security risk to an acceptably low level. his paper reviews different kinds of intrusion detection methods based on IDS performance and its effect in a cloud environment.

An effective intrusion detection system is obtained according to other types of IDS methodologies, including traditional IDS techniques such as signature-based IDS, anomaly-based IDS, host-based intrusion detection system, network-based intrusion detection system, hypervisor-based intrusion detection system, distributed based intrusion detection system, and machine learning-based intrusion detection system.

Literature Survey

Aljehane et al. (2024) proposed a hybrid intrusion detection framework that leverages the Golden Jackal Optimization Algorithm (GJOA) in combination with deep learning for improved network security. The model uses GJOA for feature selection, which enhances the performance of deep learning classifiers, achieving higher detection accuracy and computational efficiency compared to conventional optimization approaches [35].

Mahmood et al. (2024) introduced a model that integrates machine learning with multi-factor authentication (MFA) to bolster network intrusion detection. Their system combines supervised classifiers with layered authentication to prevent unauthorized access, significantly improving resilience against insider threats and identity spoofing [36].

Zhukabayeva et al. (2024) developed a machine learning-based framework for intrusion detection in wireless sensor networks (WSNs) within smart grid environments. Their approach uses traffic analysis and node categorization to dynamically select appropriate classifiers, thereby optimizing detection accuracy and

reducing energy consumption in sensor nodes.[37]

Zhao et al. (2024) applied deep learning methods, including convolutional and recurrent neural networks, to develop an intrusion detection system capable of identifying anomalies in real-time network traffic. Their system is particularly effective in detecting encrypted or zero-day threats, offering a more adaptive and low-latency solution than traditional machine learning methods [38].

Ciric et al. (2024) proposed a modular deep learning-based IDS architecture tailored for simulating and defending against real-world cyber-attacks. The system's modularity supports scalable deployment and achieves high accuracy across multiple threat types using synthetic and real datasets [39].

Ioannou et al. (2024) presented GEMLIDS-MIoT, a federated learning-based intrusion detection system designed for Medical Internet of Things (MIoT) networks. Emphasizing energy efficiency and data privacy, the model distributes learning across devices, reducing centralized communication while maintaining strong detection performance [40].

SaiSindhuTheja et al. (2021) proposed an intrusion detection system using the OCSA (Optimized Cyber Security Architecture) and Recurrent Neural Networks (RNN), specifically aimed at detecting Denial of Service (DoS) attacks. Their method proved efficient in identifying such attacks; however, it lacked a robust attack prevention mechanism, limiting its real-world applicability. Similarly, Mondal et al. (2021) utilized a hypervisor-based IDS within an OpenStack environment to detect DoS attacks, offering a fast and effective detection approach. Nonetheless, the system suffered from high memory consumption, which impacts scalability.[17]

Krishnaveni et al. (2020) employed a Support Vector Machine (SVM) model for DoS attack detection. This approach achieved higher accuracy and minimized the false alarm rate, but its deployment in complex and dynamic environments remains challenging. Taj et al. (2020) developed an Agent-Based Intrusion Detection System (ABIDS) to detect a wide range of attacks including Shellcode, DDoS, BackDoor, DoS, Fuzzers, and Reconnaissance. While the system is computationally powerful and cost-effective, it is resource-intensive and requires more processing time.

Ribeiro et al. (2020) introduced HIDROID, a system designed to detect zero-day attacks efficiently and cost-effectively. Despite its promising performance, it required further optimization to enhance its stability and efficiency [20]. Osamor et al. (2020) proposed a hybrid anomaly detection model focused on DDoS attacks. The model offered improved security but was constrained by difficulties in fully exploiting underlying vulnerabilities.[21]

Aldribi et al. (2020) worked on a hypervisor-based cloud IDS for insider attacks, showing improved detection capabilities. However, its poor resource utilization posed a limitation for practical cloud environments. Almomani et al. (2020) introduced the OIDCS (Optimal Intrusion Detection and Classification System), targeting U2R, R2L, Probe, DoS, and zero-day attacks. It demonstrated high accuracy but was computationally expensive in both cost and time.

Hajimirzaei et al. (2019) combined neural networks with the Artificial Bee Colony (ABC) algorithm to detect DoS, Probe, R2L, and U2R attacks. Their system minimized incorrect classifications but suffered from increased execution time.[24]. Kanimozhi et al. (2019) applied a multi-layer perceptron for botnet attack detection, achieving high accuracy but relying on a relatively weak security model. Abusitta et al. (2019) leveraged deep learning techniques to detect normal and malicious behaviors across U2R, R2L, Probe, and DoS attack types. The approach improved cooperative intrusion detection accuracy but introduced undesired system delays. Dey et al. (2019) implemented a KD (K-distance) algorithm to tackle DDoS and Man-In-The-Middle (MITM) attacks, delivering an efficient and secure cognitive IDS, albeit one constrained by low-speed network traffic handling.[26]

Niazi et al. (2019) proposed a Bayesian game-theoretic IDS for DDoS attack detection, notable for its cost efficiency and resource management. However, the system consumed more energy than ideal. Manickam et al. (2019) used PFCM (Possibilistic Fuzzy C-Means) and RNN for DoS detection, achieving high classification accuracy, though the method was time-consuming.[28]

Aloqaily et al. (2019) designed an automated, secure, continuous cloud service framework targeting R2L, U2R, probing, and DoS attacks. While it improved communication among vehicular nodes, it lacked integration with real-world vehicle traffic safety measures.[30]. Patil et al. (2019) introduced the HLDNS framework to detect multiple attack types including Shellcode, Worms, Fuzzers, DoS, Backdoor, and Reconnaissance. Although it minimized computation cost, it lacked secure encrypted data integration within the IDS.

Besharati et al. (2019) created LR-HIDS for detecting various attacks such as U2R, R2L, DoS, and Probe, achieving 97.51% efficiency using the NSL-KDD dataset. However, it struggled with scalability when faced with large and complex datasets.[32]. Manickam et al. (2019) also introduced a GSO–TS algorithm for DoS detection that had a lower average detection rate and suffered from high computational complexity. Ghosh et al. (2019) proposed the CS-PSO method to handle all types of attacks with rapid classification capabilities, but it demanded high memory usage. Finally, Colom et al. (2018) developed a scheduling framework for DoS attacks that effectively avoided single points of failure, though it incurred high energy consumption and required substantial memory resources.[34].

The paper explores intrusion detection in cyber security environments, employing various techniques such as signature-based, anomaly-based, host-based, network-based, hypervisor-level, distributed, and machine learning-based IDS models. However, several limitations and research gaps persist in adapting these methods to the cyber security infrastructure.

Limitations and Research Gaps

1. **False Positives and Alert Fatigue**: IDS often generate false positives, where legitimate activities are mistakenly flagged as threats. This leads to overwhelming numbers of alerts that can drain security teams' resources and cause alert fatigue,

reducing the ability to respond promptly to real threats.

2. **Limited Visibility into Encrypted Traffic**: Encrypted network traffic poses a challenge for IDS, as they cannot inspect the encrypted payload without decryption. This creates blind spots that attackers can exploit, hindering threat detection.
3. **Resource Intensive**: IDS require substantial computing power and bandwidth to analyze traffic in real-time, potentially impacting network performance and demanding high-end hardware and expert personnel. This can be a limitation for organizations with constrained resources.
4. **Passive Nature**: Traditional IDS are passive monitoring tools that detect and alert on threats but do not take active measures to block or prevent attacks, necessitating manual intervention for mitigation.
5. **Complexity and Scalability Challenges**: Deploying and managing IDS effectively in large or complex environments is challenging. It needs to be carefully adjusted, rules must be managed, and other security technologies must be integrated. Scalability and maintaining performance across distributed networks can become difficult.
6. **Signature Evasion by Attackers**: Attackers can evade IDS detection by using new or unknown attack methods that do not match existing signatures, making signature-based detection less effective against sophisticated or novel threats
7. **Automated and Efficient Feature Selection Tools**: Need for methods to automate feature extraction and selection to improve IDS accuracy and reduce manual intervention.
8. **Handling Big Data and High-Speed Networks:** Exploring scalable data mining techniques to efficiently process increasing data volumes without performance degradation.
9. **Improved Models for Encrypted Traffic**: Gaps in analyzing encrypted network traffic without decryption to enhance IDS capabilities.
10. **Integration of Multi-source Data**: Research needed on combining diverse data types (network, host, application logs) to improve the accuracy of intrusion detection.
11. **Explainability and Interpretability of Models:** Gap in providing understandable explanations for IDS alerts generated by complex machine learning models.

Addressing these challenges through innovative research will enhance the accuracy, efficiency, and practical applicability of IDS, thereby strengthening cybersecurity defenses in increasingly complex and data-rich environments.

Conclusion and Future Enhancement

This article delves deeply into the existing research on cybersecurity and intrusion detection systems (IDS) that rely on data mining, particularly in cloud and network settings. It highlights how these systems tackle various types of attacks. The findings reveal that several IDS models—like

signature-based, anomaly-based, host-based, and network-based IDS—have been thoroughly investigated. The article also delves deeper into specialized versions, such as hypervisor-based IDS, distributed IDS, and those enhanced by machine learning. A detailed survey was carried out, cataloguing the features of each IDS type, including the methods used, data mining or machine learning techniques applied, the attack vectors they target, and their respective strengths and weaknesses. There's a strong focus on data-driven strategies that incorporate classification, clustering, and anomaly detection methods into IDS frameworks to proactively spot cyber threats. Moreover, the literature reviewed points out that while traditional IDS methods are good at catching known attacks, they often fall short against zero-day exploits, polymorphic malware, and large-scale distributed attacks. Machine learning and data mining are proving to be game-changers, helping to automatically identify patterns in complex cybersecurity datasets, which allows systems for detecting which are scalable and flexible. Lastly, the analysis critically examines the current literature to pinpoint unresolved issues and research gaps, such as the high false positive rates in anomaly-based systems, the challenges of managing big data in cybersecurity, the limited interpretability of deep learning models, and the struggle to adapt to changing attack patterns.

This analysis lays the groundwork for encouraging further research into developing hybrid, data mining-assisted, and machine learning-driven IDS frameworks that can bolster resilience in cybersecurity environments.

References:

1. Sadeeq, M. M., Abdulkareem, N. M., Zeebaree, S. R., Ahmed, D. M., Sami, A. S., & Zebari, R. R. (2021). IoT and cloud computing issues, challenges and opportunities: A review. *Qubahan Academic Journal*, *1*(2), 1-7.
2. Alouffi, B., Hasnain, M., Alharbi, A., Alosaimi, W., Alyami, H., & Ayaz, M. (2021). A systematic literature review on cloud computing security: Threats and mitigation strategies. *IEEE Access*, *9*, 57792-57807.
3. Bello, S. A., Oyedele, L. O., Akinade, O. O., Bilal, M., Delgado, J. M. D., Akanbi, L. A., ... & Owolabi, H. A. (2021). Cloud computing in construction industry: Use cases, benefits and challenges. *Automation in Construction*, *122*, 103441.
4. Chen, Y., Chen, X., Liu, W., Zhou, Y., Zomaya, A. Y., Ranjan, R., & Hu, S. (2020). Stochastic scheduling for variation-aware virtual machine placement in a cloud computing CPS. *Future Generation Computer Systems*, *105*, 779-788.
5. Priyadarshini, R., Quadir Md, A., Rajendran, N., Neelanarayanan, V.,

& Sabireen, H. (2022). An enhanced encryption-based security framework in the CPS cloud. *Journal of Cloud Computing*, *11*(1), 64.

6. Xu, Z., Zhang, Y., Li, H., Yang, W., & Qi, Q. (2020). Dynamic resource provisioning for cyber-physical systems in cloud-fog-edge computing. *Journal of Cloud Computing*, *9*(1), 1-16.
7. Vinoth, S., Vemula, H. L., Haralayya, B., Mamgain, P., Hasan, M. F., & Naved, M. (2022). Application of cloud computing in banking and e-commerce and related security threats. *Materials Today: Proceedings*, *51*, 2172-2175.
8. Alam, T. (2020). Cloud computing and its role in the information technology. *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, *1*(2), 108-115.
9. Kushwah, G. S., & Ranga, V. (2021). Optimized extreme learning machine for detecting DDoS attacks in cloud computing. *Computers & Security*, *105*, 102260.
10. Devi, B. T., Shitharth, S., & Jabbar, M. A. (2020, March). An appraisal over intrusion detection systems in cloud computing security attacks. In *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 722-727). IEEE.
11. Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J., & Ahmad, F. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies*, *32*(1), e4150.
12. Kasongo, S. M., & Sun, Y. (2020). A deep learning method with wrapper based feature extraction for wireless intrusion detection system. *Computers & Security*, *92*, 101752.
13. Aldweesh, A., Derhab, A., & Emam, A. Z. (2020). Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. *Knowledge-Based Systems*, *189*, 105124.
14. Hady, A. A., Ghubaish, A., Salman, T., Unal, D., & Jain, R. (2020). Intrusion detection system for healthcare systems using medical and network data: A comparison study. *IEEE Access*, *8*, 106576-106584.

15. Mendonça, R. V., Teodoro, A. A., Rosa, R. L., Saadi, M., Melgarejo, D. C., Nardelli, P. H., & Rodríguez, D. Z. (2021). Intrusion detection system based on fast hierarchical deep convolutional neural network. *IEEE Access*, *9*, 61024-61034.
16. SaiSindhu Theja, R., & Gopal K Shyam. (2021). An efficient metaheuristic algorithm based feature selection and recurrent neural network for DoS attack detection in cloud computing environment. *Applied Soft Computing*, *100*, 106997.
17. Mondal, S., & Choudhary, A. (2021). Combating DoS attack on OpenStack using hypervisor based intrusion detection system with the help of machine learning. In *Proceedings of International Conference on Big Data, Machine Learning and their Applications* (pp. 161-168). Springer, Singapore.
18. Krishnaveni, S., Palani Vigneshwar, S., Kishore, B., Jothi & Sivamohan, S. (2020). Anomaly-based intrusion detection system using support vector machine. In *Artificial Intelligence and Evolutionary Computations in Engineering Systems* (pp. 723-731). Springer, Singapore.
19. Taj, M. S., Ullah, S. I., Salam, A., & Khan, W. U. (2020). Enhancing anomaly based intrusion detection techniques for virtualization in cloud computing using machine learning. *International Journal of Computer Science Information Security*, *18*(5).
20. Ribeiro, J., Saghezchi, F. B., Mantas, G., Rodriguez, J., & Abd-Alhameed, R. A. (2020). Hidroid: prototyping a behavioral host-based intrusion detection and prevention system for android. *IEEE Access*, *8*, 23154-23168.
21. Osamor, F., & Girma, A. (2020). Hypervisor based IDS solution approach using hybrid anomaly detection model in cloud computing environment. In *Proceedings of the Future Technologies Conference* (pp. 909-920). Springer, Cham.
22. Aldribi, A., Traore, I., Moa, B., & Nwamuo, O. (2020). Hypervisor-based cloud intrusion detection through online multivariate statistical change tracking. *Computers & Security*, *88*, 101646.

23. Almomani, A., Alauthman, M., Albalas, F., Dorgham, O., & Obeidat, A. (2020). An online intrusion detection system to cloud computing based on neuCube algorithms. In *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications* (pp. 1042-1059). IGI Global.
24. Hajimirzaei, B., & Navimipour, N. J. (2019). Intrusion detection for cloud computing using neural networks and artificial bee colony optimization algorithm. *ICT Express*, *5*(1), 56-59.
25. Kanimozhi, V., & Prem Jacob, T. (2018). Artificial intelligence based network intrusion detection with hyper-parameter optimization tuning on the realistic cyber dataset CSE-CIC-IDS using cloud computing. In *International Conference on Communication and Signal Processing (ICCSP)* (pp. 33-36). IEEE.
26. Abusitta, A., Bellaiche, M., Dagenais, M., & Halabi, T. (2019). A deep learning approach for proactive multi-cloud cooperative intrusion detection system. *Future Generation Computer Systems*, *98*, 308-318.
27. Dey, S., Ye, Q., & Sampalli, S. (2019). A machine learning based intrusion detection scheme for data fusion in mobile clouds involving heterogeneous client networks. *Information Fusion*, *49*, 205-215.
28. Niazi, R. A., & Faheem, Y. (2019). A Bayesian game theoretic intrusion detection system for hypervisor-based software defined networks in smart grids. *IEEE Access*, *7*, 88656-88672.
29. Manickam, M., & Rajagopalan, S. P. (2019). A hybrid multi-layer intrusion detection system in cloud. *Cluster Computing*, *22*(2), 3961-3969.
30. Aloqaily, M., Otoum, S., Al Ridhawi, I., & Jararweh, Y. (2019). An intrusion detection system for connected vehicles in smart cities. *Ad Hoc Networks*, *90*, 101842.
31. Patil, R., Dudeja, H., & Modi, C. (2019). Designing an efficient security framework for detecting intrusions in virtual network of cloud computing. *Computers & Security*, *85*, 402-422.
32. Besharati, E., Naderan, M., & Namjoo, E. (2019). LR-HIDS: logistic regression host-based

intrusion detection system for cloud environments. *Journal of Ambient Intelligence and Humanized Computing*.

33. Ghosh, P., Karmakar, A., Sharma, J., & Phadikar, S. (2019). CS-PSO based intrusion detection system in cloud environment. In *Emerging Technologies in Data Mining and Information Security* (pp. 261-269). Springer, Singapore.
34. Colom, J. F., Gil, D., Mora, H., Volckaert, B., & Jimeno, A. M. (2018). Scheduling framework for distributed intrusion detection systems over heterogeneous network architectures. *Journal of Network and Computer Applications*, *108*, 76-86.
35. Aljehane, N. O., Mengash, H. A., Eltahir, M. M., Alotaibi, F. A., Aljameel, S. S., Yafoz, A., ... & Assiri, M. (2024). Golden jackal optimization algorithm with deep learning assisted intrusion detection system for network security. *Alexandria Engineering Journal*, *86*, 415-424.
36. Mahmood, R. K., Mahameed, A. I., Lateef, N. Q., Jasim, H. M., Radhi, A. D., Ahmed, S. R., & Tupe-Waghmare, P. (2024). Optimizing network security with machine learning and multi-factor authentication for enhanced intrusion detection. *Journal of Robotics and Control (JRC)*, *5*(5), 1502-1524.
37. Zhukabayeva, T., Pervez, A., Mardenov, Y., Othman, M., Karabayev, N., & Ahmad, Z. (2024). A traffic analysis and node categorization-aware machine learning-integrated framework for cybersecurity intrusion detection and prevention of WSNs in smart grids. *IEEE Access*, *12*, 91715-91733.
38. Zhao, F., Li, H., Niu, K., Shi, J., & Song, R. (2024). Application of deep learning-based intrusion detection system (IDS) in network anomaly traffic detection.
39. Ciric, V., Milosevic, M., Sokolovic, D., & Milentijevic, I. (2024). Modular deep learning-based network intrusion detection architecture for real-world cyber-attack simulation. *Simulation Modelling Practice and Theory*, *133*, 102916.
40. Ioannou, I., Nagaradjane, P., Angin, P., Balasubramanian, P., Kavitha, K. J., Murugan, P., & Vassiliou, V. (2024). Gemlids-miot: A green effective machine learning intrusion

detection system based on federated learning for medical IoT network security hardening. *Computer Communications, 218*, 209-239.