STAP
Smart Technologies Academic Press

STAP Journal of Security Risk Management

https://www.jsrm.thestap.com/

Enhancing Intrusion Detection Systems by Using Machine Learning in Smart Cities: Issues, Challenges and Future Research Direction

Rasha Almarshood[1], M. M. Hafizur Rahman[1]

[1] *Department of Computer Networks and Communications, College of Computer Sciences and Information Technology, King Faisal University, Al-Ahsa 31982, Saudi Arabia*

ARTICLE INFO

Article History

Received: 24-03-2025
Revised: 20-05-2025
Accepted: 22-05-2025
Published: 25-05-2025

Vol.2025, No.1

DOI:

***Corresponding author.**
Email:
mhrahman@kfu.edu.sa

Orcid:
https://orcid.org/0000-0001-6808-3373

ABSTRACT

With promising innovation and efficiency in smart city, it is still facing a growing threat of cyberattacks. The increasing interconnectedness of digital services makes these cities particularly vulnerable. Traditional security measures struggle to adapt to evolving threats. Due to the insufficient analysis of real-time attack patterns. Emerging new technologies are crucial for managing these issues. Machine Learning (ML) is a promising solution to enhance Intrusion Detection Systems (IDS). ML can effectively detect malicious activities. ML provides automation of network traffic analysis and anomalous pattern identification. This paper presents a systematic literature review to explore the potential of ML in improving IDS for smart city. Various ML approaches and specific applications in smart city services will be investigated. We will evaluate the effectiveness of existing approaches in smart city. Identifying key challenges and future research directions. We also aim to contribute to the development of smart city security systems. It will benefit critical infrastructures to be more robust and resilient against evolving threats.

Keywords: IDS, ML, Smart City, Anomaly Detection, Real-time Analysis.

How to cite the article

1. Introduction

Smart cities have become more convenient nowadays. Due to the rapid growth of technologies that improve the quality of life. Many countries have adopted the idea of transforming into smart cities. By providing sustainable, secure, and consistent services without disruption. Connectivity is the most important aspect of these environments. All smart city applications need to be connected to be managed and processed instantly. Smart grid, smart agriculture, smart healthcare, and smart transportation are examples of smart city applications [1]. It is necessary to secure these systems to ensure that the privacy of public or private information is not compromised. With the increasing number of cyber threats, systems cannot always be secured.

DDoS, data breaches, zero-day attacks, spoofing, and eavesdropping are examples of common cybersecurity [2]. To prevent such threats, it is important to implement strong defense strategies such as IDS. IDS is responsible for monitoring network traffic and identifying unauthorized behaviors. IDS can monitor the data exchange of smart city systems and devices. However, protecting and securing big data against new attacks requires a fast response in real time. While IDS can detect threats predefined by rules and signatures. Therefore, integration of new technology, such as ML, is crucial [3].

ML has played a significant role in enhancing smart city applications. For instance, ML has been used to analyze big data in industries such as healthcare systems and government sectors. It improves the quality of services that smart cities need. ML uses algorithms that can analyze pre-given data to identify different patterns of suspicious behaviors. ML has improved the IDS features by providing an approach that can early predict and alert of a cyber threat instantly. This allows stakeholders to take appropriate countermeasures to prevent an attack. Integration of ML with IDS will improve the quality of services provided and will ensure a secure system of smart cities [4].

The main aim is to improve the detection and prediction system against cyberattacks. IDS can be enhanced by integrating ML approaches. Due to the evolving cyberattacks, a robust framework is needed. There are specific vulnerabilities related to critical industries. Such as weak network configurations and lack of real-time data analysis. These vulnerabilities can expose industrial systems to critical situations. A proactive detection system that analyzes big data is essential. This will provide the enhancement of the IDS capabilities. Additionally, implementing approaches such as Artificial Neural Networks (ANN) and Support Vector Machine (SVM). These approaches are used to deviate malicious activities from normal data traffic. By addressing these gaps, evolving threats and zero-day attacks will be reduced and mitigated [5].

In this paper, we will highlight the importance of using IDS in smart city industries. Explore the types of IDS. Highlight the significant role of emerging ML with IDS. Exploring the ML applications on IDS. And highlight the cyber-attacks related to IDS. Address the limitations that have been identified. Recommendations will be addressed to provide resilience and reliability in security systems. The objective of this paper can be defined as follows:

- Review and analyze the existing literature regarding the emerging ML with IDS.
- Investigate current ML algorithm limitations.
- Highlight the enhancement of IDS-based ML.
- Provide recommendations to improve IDS-based ML in smart cities.

This paper is structured as follows: Section 2 methodology for selecting related studies. In section 3 provides a background on IDS and its classifications. Section 4 highlights the significant emerging of ML in IDS. Discusses the cyber-attacks related to IDS. Section 5 highlight the process of analyzing big data and the significant role of ML. Section 6 review related studies and compare previous research to highlight the limitations and contributions. Section 7 Discussion. Section 8 discuss open challenges and future recommendations. 9 Conclusion.

2. Literature Review

This paper has been conducted for a comprehensive analysis on the integration of ML in IDS in smart city applications. In this study, we have used a PRISMA flow diagram for creating our systematic review to select and analyze the papers that are relevant to our study. Systematic literature review (SLR) aims to write research by locating, picking, and critically evaluating all findings from all papers. Our paper aims to identify the gaps in the current studies and suggest future enhancements in this field. This SLR contains four main stages, which are identification, screening, eligibility, and inclusion. The focus of the literature review was the published papers between 2020 and 2024. We chose academic journals or

conferences as the source type using a search filter. We determined the search strings and identified the relevant sources of data. Studies were selected based on specific inclusion and exclusion criteria. These criteria are important for ensuring that the research findings and data are relevant, reliable, and applicable to the research.

2.1. Search String

Search strings, data sources, and keywords were identified. And studies were selected based on criteria. Our search string contains of Boolean operators such as “AND”, “OR” between the keywords. These operators will be very helpful in the widening, narrowing, and refining of the search string.

2.2. Data Sources

The SLR is focused on papers published between 2020 and 2024 from academic sources.In this paper, we collected our papers by using databases such as Google Scholar, SinceDirect, MDPI, IEEE, and the Saudi Digital Library.

2.3. Screening Process

We have specified the search string to ensure that their titles include the words “IDS”, “ML”, and “smart cities”. The duplicate results will be removed from our search string. The next step is the screening process, which excludes the irrelevant papers for our topic. Then, we will assess the current result for eligibility and exclude the irrelevant papers. Finally, we will include the final set of papers in our literature review. The published year for the selected papers was between 2020 and 2024. The PRISMA methodology used in our paper is illustrated in Figure 1.

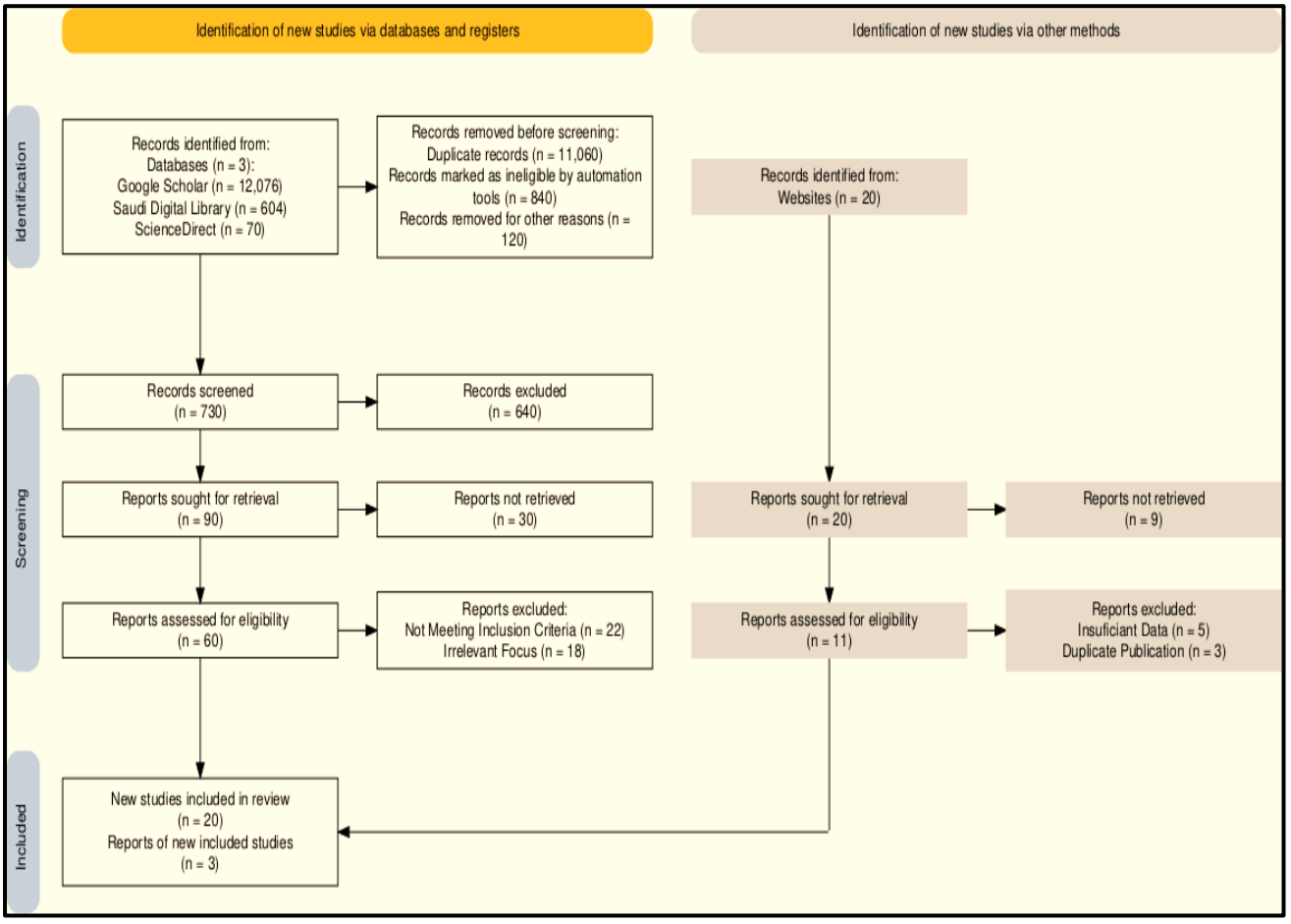

Figure 1. Selection of Research Papers for Literature Review Using PRISMA 2020 Model.

3. Intrusion Detection System in Smart City

Smart city depend on a large network of interconnected systems. These systems are targeted to a variety of cybersecurity threats. IDS provides a powerful solution to proactively identify and respond to potential threats. When network traffic and system logs are analyzed, IDS will detect and prevent security threats. Following are the roles of IDS applications in enhancing smart city services.

3.1. Monitoring Network Traffic

IDS play a significant role in protecting smart city in diverse services [6]. One of these services is a continuous network traffic monitor. This monitoring will help to identify signs of malicious attacks. IDS use a combination of techniques to identify potential threats [7]. These techniques are signature-based detection, anomaly detection, and behavioral analysis [8]. Common threats related to IDS are DoS attacks, data breaches, unauthorized access attempts, and evolving threats. When an attack is detected and alerted, IDS enables administrators to take proactive measures [4].

3.2. Identifying Security Breaches

IDS has a significant role in protecting sensitive data for critical infrastructures. Analyzing network traffic and system logs will detect a variety of security threats. For instance, data breaches can compromise sensitive data. DDoS attacks may disrupt vital services functioning in smart city. IDS will identify these threats and help to ensure security protection and resilience for smart city.

3.3. Enhancing Cyber Security

Additionally, IDS enhance the cybersecurity of critical infrastructures. IDS will provide early alarms of potential threats. It will continually monitor network traffic and detect malicious activities in real time. This will enable security administrators to respond proactively and mitigate risks. Those proactive measures will ensure resilience and security.

3.4. Critical Infrastructure Protection

Smart city is relying on a variety of critical infrastructures. Such as healthcare, transportation, and smart grids. These systems are vulnerable to security threats due to their interconnection. Therefore, IDS provides protection for these infrastructures by alerting and detecting security threats. It will proactively identify and respond to potential threats. IDS will ensure the reliability and security of vital services. By reducing disruptions and providing protections of smart city services.

3.5. Real-time Response and Mitigation

For real-time response and mitigation, IDS plays a significant role in this field. It initiates specified actions automatically. For instance, restricting malicious activities and blocking compromised systems from others. It also prevents attacks and reduces the impact on critical systems. IDS enabling incident response by using forensic analysis. As well as enhancing future security measures. This proactive method will ensure security protection for critical services in smart city.

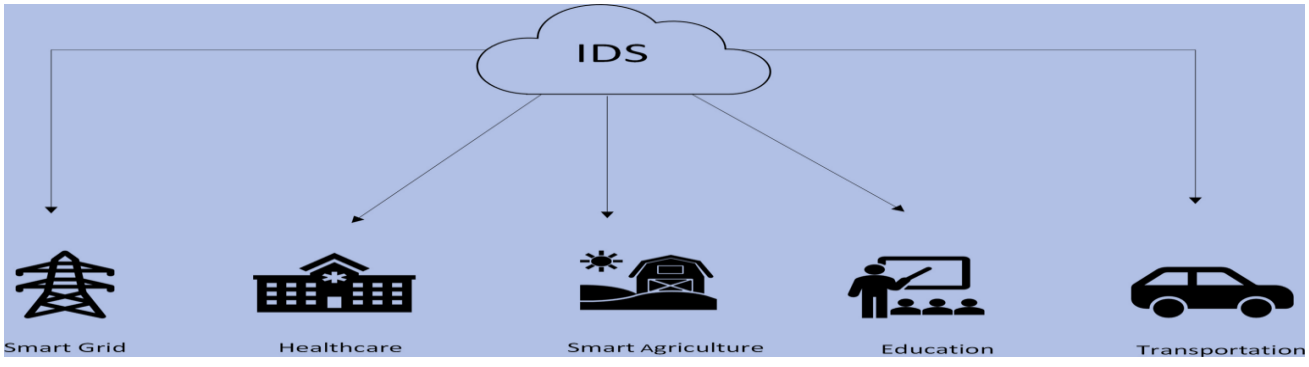

Figure 2. IDS Applications in Smart City

3.6. Evolution of IDS

IDS has developed significantly over the years. The old version of IDS relied on predefined rules. However, the evolved ids use the latest technologies. For instance, ML and AI prove their effectiveness in detecting threats. They have evolved from detecting threats to becoming the most important tools to prevent and respond to any security threats. The latest trends in modern IDS are cloud-based IDS, behavior-based detection, integration of IDS with other security tools, and automation and orchestration. Cloud-based IDS: to enhance the performance of anomaly and signature-based IDS by using approaches to robustly secure the security of cloud environments [9]. Behavior-based IDS helps to detect the abnormal behaviors that lead to cyber-attacks [10]. Integration IDS in security tools, it evolved the performance of these tools. IDS have been used in the automation and orchestration of distributed energy resources (DER) in effective ways [11]. Therefore, IDS is considered one of the robust security tools that can enhance the security of organizations against cyber-attacks. Figure 3 explains the integration of ML in IDS.

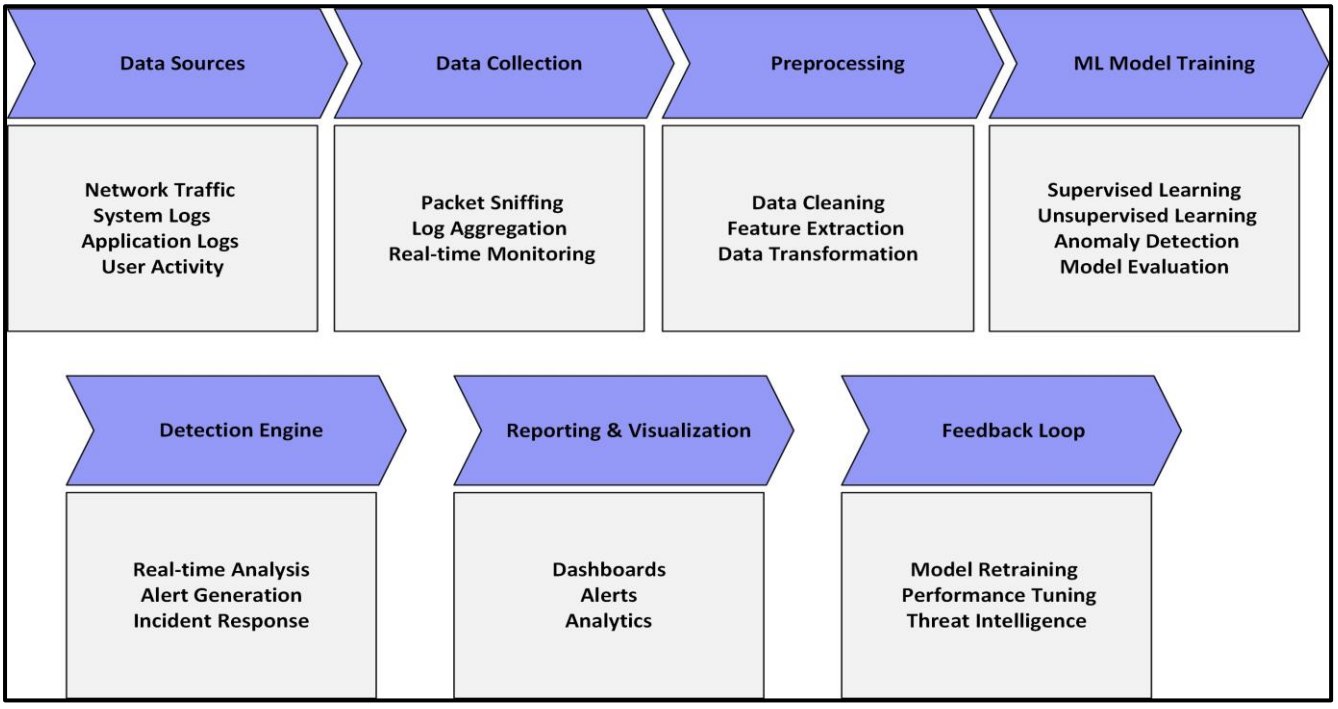

Figure 3. ML Integrated with IDS

3.6.1. Machine Learning Applications in IDS

Integration of machine learning in intrusion detection systems has significantly evolved the capabilities by enhancing the effectiveness of detecting and responding to cyber threats. ML techniques in IDS have developed the capability of reducing the false positive alarm, developing the threat landscape, and taking proper security measures. Following are some approaches of ML in IDS:

• Supervised Learning Techniques Supervised learning techniques are commonly applied in IDS. It also provides labeled datasets for training purposes. These labelled datasets are used to predict and classify security threats. In cybersecurity domains, classification and regression approaches are commonly used as supervised learning tools. According to [12], decision trees are considered one of the most significant and important tools for predicting and classifying threats. Support victim machine (SVM), also used for classifying and predicting by finding the optimal hyperplane to separate different classes in high-dimensional space [8]. Neural networks can identify the suspicious patterns through multiple layers [13]. The result of multiple combined algorithms provided the most accurate and effective anomaly detection [14]. Applying the proper techniques of supervised learning approaches will effectively evolve the IDS security capabilities that stockholders are looking for.

• Unsupervised Learning Techniques detecting anomaly patterns by IDS using unsupervised learning techniques is crucial. These techniques can effectively enable IDS to adapt to the devolving threats. Furthermore, these techniques don't require any labeled datasets for training that can ensure effective response and detection methods. K-Means clustering is an example of a clustering algorithm that splits data-based similarity into K distinct clusters. It identifies and detects the similarity of behaviors using the clustering method to address these issues [12]. Another method of unsupervised learning technique is anomaly detection. Isolation forest is an algorithm of anomaly detection that identifies the anomalies analyzed. It works based on the rarity of anomaly data and gets isolated in a minimum number of portions [15]. Isolation forest depends on randomized special i tree values. According to [16], isolation forest has two phases, which are the generation of randomly i trees. And assign a score for each database observed.

• Reinforcement Learning Techniques Reinforcement learning (RL), integrated in IDS, can be adequate with recent security changes and evolve at the same time based on the interactions with the environment. The Q-Learning model is one of the reinforcement learning techniques that has been used in IDs. The model evaluate the different actions, taking into consideration the current situation of the agent [17]. Multi-agent reinforcement learning (MARL) is the interaction of multiple agents with their environment.

An agent will be responsible for a particular policy by taking into consideration the actions of other agents [18]. It requires the collaboration of all agents to work together, which makes it more challenging compared to independent single-agent reinforcement learning. Therefore, it's crucial to choose the effective configuration and structure of the model that will provide a powerful framework against cyber threats. The policy gradients model is very promising due to its ease of implementation without any settings required. It can adapt to different policy parameters in large environments [19]. Entities can ensure robust security measures by identifying policies and enforcing them effectively.

4. ML-based IDS Applications in Smart City

Implementation of machine learning techniques in ids has evolved the services of smart cities efficiently. A variety of these approaches provide successful detection and real-time response against cyber threats. By ensuring proactive resolutions to safeguard smart city environments. Following are some case studies of the integration of ML-based IDS in smart cities. A successful approach was adopted to detect anomalies in traffic data to monitor traffic activities in this paper [20]. The main goal of this article was to detect the anomalies in traffic data. Such as traffic flow, accidents, and traffic jams. Using this technique has enhanced traffic system management and optimization. The paper has discussed the integration of machine learning methods by installing sensors on the side roads to analyze the collected data to identify the abnormal traffic patterns. These data were analyzed using techniques such as anomaly detection techniques and clustering algorithms. Another approach that was integrated into the healthcare system to detect abnormal activities in medical images was introduced in this paper [5]. This approach was trained to learn a wide range of healthy medical images on large datasets. If a new medical image is uploaded, the approach can detect the abnormality of this input by comparing it with what the database was trained for. This paper [21] discussed the approach used in surveillance footage. By training the model on large datasets of normal surveillance footage. The approach will detect any suspicious or unusual activities that have occurred. By comparison between the new analyzed footage and the trained datasets to show the difference between them. A convolutional neural network can be used in such a model to train on large databases of labeled data to identify the normal activities. Such an approach has enhanced the ability to understand the behaviors in surveillance footage in smart cities. If technology is continuously evolving, a variety of effective applications in different fields will be observed.

4.1. Cybersecurity Challenges and Mitigations

Evolving smart city industries require a robust security system that prevents potential attacks. Any vulnerabilities in critical systems will cause serious damage. Therefore, analyzing real-time data is crucial to identifying potential threats. Mitigate and respond to identified threats. And ensuring that smart city industries security systems are secure and resilient. Following are some cybersecurity challenges related to IDS. Figure 4 shows the security challenges related to IDS.

4.1.1. Evolving Threats

The landscape of cybersecurity is evolving continuously. Malware and phishing attacks are commonly known. However, new attacks have been observed, such as supply chain attacks and ransomware. IoT devices can be affected by these potential

threats. Attackers can launch attacks by compromising ML algorithms. It will ultimately cause a big system disruption. To mitigate these threats, it is recommended to implement developed IDS frameworks. Ensuring to detect and respond to zero-day attacks and evolving threats effectively.

4.1.2. IDS Configurations

Misconfiguration of the IDS leads to serious incidents. Only security experts are authorized to monitor and deploy IDS. Any weakness in the configurations will make them vulnerable to security threats. Security experts cannot handle and be proactive to a huge number of alerts that are generated by IDS. To enhance and mitigate this vulnerability, it's important to configure the IDS according to their network requirements. Therefore, smart city industries must deploy threat intelligence sharing platforms. This will help to be up to date with cybersecurity trends. Maintain and ensure that IDS is updated to effectively mitigate potential threats.

4.1.3. Distributed Denial of Service Attack (DDoS)

A DDoS attack is one of the common security risks that has a direct impact on the availability of the data. It simply sends a flood of requests to the target machine to cause the system to outrun the service. Smart city depends on sensors and automation for controlling, managing, and monitoring. A DDoS attack has the ability to shut down the system. Moreover, such attacks are affecting the privacy of critical data and data harm in smart cities. To overcome this attack, an intrusion detection system can help to look for and alert of illegitimate or suspicious behavior on the network. Also, ML has the ability to analyze an enormous amount of data and deploy it to check up on normal behaviors.

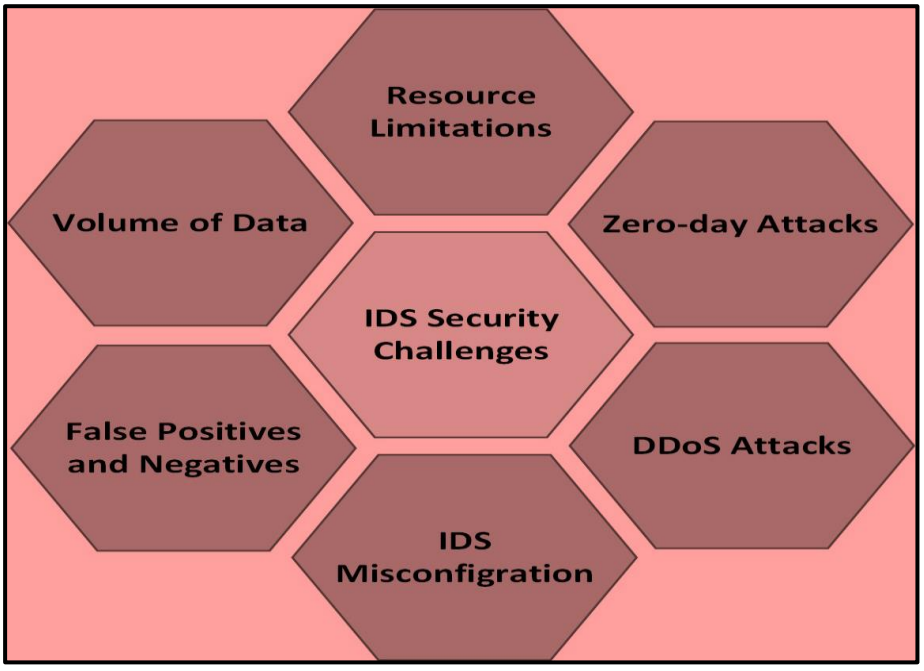

Figure 4. IDS Security Challenges

5. Analyzing Big Data with ML

Analysis of big data requires significant efforts that humans cannot do. Even if experts are analyzing the data, it will take time and efforts that smart cities cannot wait. Not to mention the risks related to this data if it's not analyzed properly. Therefore, ML has emerged to analyze the vast amount of the data effectively. ML played a significant role in analyzing smart city critical and regenerative data. It can enhance the overall performance by learning from previous data history. As mentioned in recent studies, ML facilitates decision-making and provides real-time insights [22]. By gathering and analyzing

data from many resources, these are essential for smart city services .This paper [1] has explained the importance and need for resilient ML approaches that process data in real time. This processing will help to identify useful information for industry management. This process will enhance the smart city services by providing clear real-time insights to foster the capabilities of these services. Current data analytics are solving multiple challenges as shown in Table 1, which will be discussed as follows. Tools like Apache Spark made large dataset management easier. It offers real-time analytics that support industries such as transportation and public safety [23]. Such a tool provides robustness and resilience enhancement for smart communities. Tableau can convert data into interactive visualization, which improves stakeholders cooperation and predictive modeling [24]. The programming language R provides time-series analysis and predictive modeling for different smart city applications. R will be a successful key in visualizing the data of smart city services in a shareable and comprehensive way [25]. Cassandra is an open-source distributed database that improves the operational efficiency of smart city organizations. It manages large data volumes and ensures resilience through its replicated architecture. Additionally, it can ensure that the system will perform effectively while maintaining the node failures. [26]. They mentioned technologies address various challenges in data analytics that improve the development of smart city services.

Table 1. Tools for Analyzing Big Data Using ML

References	Domain	Challenges	Tools	Description
[24]	Data collection and storage	Data volume and variety.	Apache Spark	Open-source unified analytics engine to process big data using built-in models such as SQL and graph processing.
[25]	Data processing and analytics	Data integration, and privacy concerns	Tableau	Visualizing tool that creates interactive dashboards for easy analytics process.
[26]	Data analysis and visualization	Complex data analysis		Programming language for statistical analysis and machine learning.
[27]	Real-time analytics	Real-time data processing needs	Cassandra	It provides real-time analysis, which helps organizations to analyze and take proper action at the same time.

6. Related Works

This section will cover the relevant studies in this field by summarizing the key findings about integrating ML in IDS in smart cities. Addressing the research limitations and gaps. And justify the possible mitigation according to the best of our knowledge presented in Table 2.

In this paper [27], the researchers developed an IDS framework that uses the XGBoost- based feature selection approach. This feature is combined with the variety of RRN architectures. It includes Long-Short Term Memory (LSTM), Gated Recurrent Unit (GRU), and simple RNN. The researchers evaluated the performance of these methods on benchmark datasets. By analyzing the effectiveness in detecting network intrusions. The paper's aim is to enhance the IDS performance. By utilizing deep learning techniques and feature selections. The challenges addressed in this paper are low detection accuracy and high false alarm rate in current IDS. The paper used more performance metrics to evaluate the methods. For instance, Total Accuracy (TAC) and False Alarm Rate (FAR). The key findings were achieving high total accuracy on the selected dataset with 87.07% effectiveness. The limitations found were the need for broad feature selection and the possibility of overfitting. Additionally, the model is being generalized to apply to different types of networks. The paper discussed the integration of ML and DL approaches in the IDS design to detect network vulnerabilities.

This paper [6] highlighted the taxonomy of ML methods by evaluating the perfor- mance of multiple IDS solutions and discussed recent trends in this field. The main objective of this paper is to improve IDS detection accuracy by reducing false alarm rates. Addition- ally, addressing the challenges of zero-day attack detection. The key findings mentioned in this paper were the significant improvement of IDS performance. Moreover, the pa- per explained how hybrid detection algorithms have promising results. The limitations highlighted were that the current IDS is still struggling with a high rate of false positives. Moreover, DL techniques have not been explored adequately.

The researchers of this paper [28] developed a data-driven approach using automated machine learning algorithms. This approach is for intrusion and anomaly detection in IoT environments. The issues addressed in this paper were the reasons for the insufficient detection systems of current IDS. Due to the classification of multi-class capabilities and ineffective accuracy. Datasets need to be balanced to enhance the performance of the model. Also, reduce computational costs and automate the selection of optimal algorithms. The paper was focused on enhancing the network security and detection capabilities of the IoT environment. The researchers proposed a model that significantly achieved 99.7% classification accuracy. Additionally, it highlighted the effectiveness of handling class imbalances that improved the detection rates. Limitations addressed were the reliability on specific datasets, which is KDDcup99. Additionally, computational resources demand to be associated with the process of auto-ML.

The paper [29] evaluated FL-based IDS in IoT systems by addressing data privacy concerns. And evaluate the successful detection of broad attacks on IoT devices. The aim of this was to identify challenges related to deploying FL in IDS for IoT environments. And evaluate the FL performance with a variety of data aggregation and techniques. This paper is focused on enhancing cybersecurity through developing intrusion detection methodologies. The key findings were that FL has the ability to enhance privacy along with detecting intrusions effectively. A variety of data distributions have significantly impacted the IDS performance. The limitations found in this paper were that many approaches are based on impractical data distribution. And the potential vulnerabilities in FL settings. In this paper [3], the researchers had developed algorithms and feature extraction methods for ML-based IDS in IoT systems. A variety of ML approaches and feature extractors have been tested and evaluated to enhance the detection accuracy rate. The aim of this paper was to address IoT device vulnerabilities, and enhance the detection of malicious activities in data traffic. The paper focused on developing an IDS model that uses feature extraction to enhance the effectiveness and accuracy of detection capabilities. The IEEE Datapor dataset has been used to evaluate the model's performance. Also, using VGG-16 as an example of feature extraction combined with ML models. The key findings were that combining feature extensions with ML approaches enhanced the accuracy of intrusion detection efficiently. Additionally, VGG-16 combined with stacking scored a high accuracy rate. The limitations found were substantial computational resources are required for deep learning algorithms. Also, the results of the presented model may differ from other types of attacks. The paper offered promising enhancements in IDS integrated in IoT environments.

In this paper [12] the researchers explored ML models integrated in cybersecurity applications. They reviewed how these models can be deployed in intelligent data analysis and automated response. The aim of the paper was to address the challenges related to traditional cybersecurity mitigations. And concerns related to the consistent evolving security threats. The paper explored the role of ML algorithms in predicting and preventing security attacks. Moreover, the paper highlighted the importance of automating the secu- rity systems. The paper focused on detecting intrusions, malware, and phishing attacks. Additionally, the predictive analytics for preventing cyberattacks have been addressed. The paper evaluated various ML algorithms to measure their effectiveness in real-world scenarios. The key findings were that the ability of ML detection and responsive capabilities has significantly improved. Moreover, automation cybersecurity systems can be efficiently deployed through ML. The limitations found were that ML is relaying on the data quality and quantity. Also vulnerabilities such as adversarial attacks were found in ML models.

The researchers of this paper [7] proposed a model that utilized ML techniques called NID-Shield. It is a hybrid network intrusion detection system (NIDS). This system can classify and analyze network data based on types and names of attacks. The aim of the paper was to enhance IDS by improving the network attack classification. Address difficulties in identifying various attacks in complex networks. They focused on IDS and network security for organizations. Additionally, they deployed CAPPER feature selection to help identify high-quality features. The key findings were that NID-Shield showed high accuracy levels and low false positive rates. These results were tested using UNSW- NB15 and NSL-KDD datasets.

The CAPPER algorithm has improved the performance of intrusion detection. The limitation was that only specific datasets were applicable for this system. It also requires further investigation to measure the adaptability of real-time attack scenarios. In this paper [30] the researchers have integrated ensemble-based ML for IDS. Additionally, they explore and evaluate different ensemble methods, such as random forest, in multiple datasets. The aim of the paper was to enhance the detection of unseen attacks. Moreover, address the gaps in current models that are dependent on specific datasets. The paper was focused on protecting computer network security through effective utilization of IDS. The key finding was that the random forest method achieved over a 99% accuracy rate among other datasets. The limitations found were that the model efficiency may differ across variety datasets. Additionally, the requirements of computing resources are limiting real-time integration. Both false positives and negatives are still occurring. However, the paper shows promising improvements that need further adaptation to address cyber-attacks.

The authors of this paper [31] used a Gina Impurity-based Weighted Random Forest (GIWRF) technique to improve the IDS model. The aim was to enhance the accuracy of detecting intrusions effectively through ML approaches. The paper highlighted the issues related to high-dimensional feature vectors in IDS. The paper is focused on IDS protection for sensitive data on networks. The key findings are the successful combination between the GIWRF and Decision Tree model. Outperformed model among others was the GIWRF- DT. The limitations of this paper are that the results depend on specific datasets and may not be applicable to other datasets. Additionally, the paper was focused only on binary classification.
In this paper [32] the researchers had developed an IoT-based IDS that aims to effectively detect security threats. By utilizing new hybrid deep learning approaches, it will be able to detect DDoS attacks in IoT systems. The aim is to analyze network traffic that used by IoT systems. By ensuring high detection accuracy, and reducing the consumption of the resources. The paper focused on IoT sector especially in wireless sensor networks. They had used Apache spark and PySpark tools with CICIoT2023 and TON_IoT datasets. The key findings was that the proposed hybrid model has achieved 99.9% accuracy rate for binary classification. Additionally, the model achieved 99.96% accuracy for multi classification. The hybrid model proved to be more effective than ML and DL algorithms. The limitations were that the paper has not address the issues of false positive and negative in attack detection. Moreover, the paper does not discuss the scalability of the applied solution in real world scenarios.

The researchers of this paper [33], trained a model that contained both normal and malicious network traffic to optimize the detection system. By using deep convolutional neural networks (DCNNs) on large datasets. The paper focused on the cyber security sector, especially in protecting computer network security. The researchers used in their study several performance metrics. Such as detection accuracy, false positive rate, and computational efficiency. The main findings are that the proposed model achieved a range between 99.79% and 100% for detection accuracy rates. It showed effectiveness compared to traditional IDS methods. The limitations are that the model is dependent on specific datasets, which may not be applicable for other types of threats. Performance may be affected due to unaddressed L1 or L2 regularization for hyperparameters.

In this paper [2] the researchers had proposed an anomaly detection mechanism that includes multiple types of ML approaches. Such as SVM, ANN, KNN, LR, DT, and RF. The paper focused on ensemble techniques to optimize the IDS performance. The aim is to address the cybersecurity threats in the IoT systems and the need for robust detection systems. The paper is focused on IoT sector in smart city services. The performance metrics used in the paper are accuracy, precision, recall, and F1score. One of the key findings is that the implementation of both multiple classifiers and feature selection has enhanced the IDS capabilities. Additionally, the proposed model has the best performance compared to other techniques. There are scalability limitations when used in complex IoT systems. Additionally, it is based on specific datasets that may minimize generalizability to realistic events.
The researchers of this paper [34], developed an IDS using DL architectures by focused to use deep neural networks (DNNs) for detection and classification. The aim was to address vulnerabilities in cybersecurity and highlight the need for advanced detection systems that is capable to detect evolving threats. The paper focused in cybersecurity sector especially in network security and IDS. The dataset used in this paper was UNSW- NB15 that simulate behaviors of modern network. Convolutional neural network was employed as DL model. And accuracy, desecrate was utilized as performance metrics the key findings was that the model has achieved remarkable accuracy rate with 95.6%. Significant improvements in multi-class classifications tasks was demonstrated in the design. A lower rate of detection was observed for underrepresented classes due to imbalance class in the datasets. The model was trained on historical data which may not adapt to newly emerging threats. Relying on specific datasets may limit the model ability to be generalized to different environments.

The researchers of this study [35] proposed a model specifically to detect threats on IoT networks. This model was designed as ML-based IDS to enhance the detection capabilities. The aim is to address security and privacy issues related to IoT devices. Also, they highlighted the limitations of IoT devices such as memory and processing power. The paper focused on IoT applications in smart cities. They applied on UNSW-NB15 dataset, and used several ML models such as XGBoost. They used several performance metrics to evaluate the model. Accuracy, area under the curve (AUC), and recall. The key findings are the model has achieved 99.9% accuracy rate. Additionally, it detect various types of IoT attacks. The proposed model may not address all types of attacks adequately. The model depends on specific datasets that may not be applicable to different environments.

The author of this paper [36] proposed a model that used four-layer deep fully con- nected (FC) network architecture. This model is DL-based IDS that used on IoT networks. It used to detect cyberattacks and malicious activity across IoT networks. The aim of the paper is to address vulnerabilities related to IoT devices. The paper focused on IoT sector, specifically in smart cities applications. The researcher used SVM model and the performance evaluation was done through realistic events. The key findings are that the proposed model achieved a 93.21% detection rate. And 93.74% in detecting different at- tacks. Additionally, the model has presented reliability in reality and simulated events. The limitations were that the model efficiency was based on the quality of the datasets. Also, the model might face adaptation challenges due to evolving threats.

Table 2. Existing Works

Reference	Key Findings	Limitations/Research Gaps	Suggested Mitigation
[27]	• The paper objective was to enhance the IDS performance by integrating deep learning techniques. • Achieving high total accuracy on the selected dataset with 87.07% effectiveness.	• The model is being generalized to apply to different types of networks. • Possibility for overfitting. • Limitations in current field.	• Implement more advanced feature selection that is capable of automatically adapting to evolving attacks. • Develop IDS framework capability to adapt to new threats by incorporating continuous learning methods.
[10]	• The significance improvement of IDS performance. • The paper explained how hybrid detection algorithms have promising results.	Highlighted was that current IDS is still struggling with a high rate of false positives. DL techniques have not been explored adequately.	• Explore more ML approaches that are able to learn complex patterns of sophisticated threats. • Use hybrid detection approaches such as misuse and anomaly detection to reduce false positive alarms.
[28]	• Proposed model that significantly achieved 99.7% of classification accuracy. • It highlighted the effectiveness of handling class imbalance that improved the detection rates.	• The reliability on specific datasets, which is KDDcup99. • Computational resources demand to associate with the process of Auto-ML.	• Use multiple datasets to expand the range of attack types and real-world scenarios. • Deploy lightweight models that reduce computational requirements. • Deploy edge computing to reduce resource demands by allowing data processing near resources.

[36]	• The proposed model achieved a 93.21% detection rate. And 93.74% in detecting different attacks. • The model has presented reliability in reality and simulated events.	• The model efficiency was based on the quality of the datasets. • The model might face adaptation challenges due to evolving threats.	• It's recommended to use continuous learning strategies to update the model with new attack data. • Gather data from different IoT environments to train the model with realistic attacks.
[29]	• FL has the ability to enhance privacy along with detecting intrusions effectively. • A variety of data distributions has significantly impacted the IDS performance.	• Many approaches are based on impractical data distribution. • The potential vulnerabilities in FL settings.	• Reduce the size of patches sent to during federated training by using model compression methods. • Enhance the learning process for new attacks by exploring federated transfer learning techniques.
[3]	• Combining feature extensions with ML approaches enhanced the accuracy of intrusion detection efficiently. • VGG-16 combined with stacking scored a high accuracy rate.	• The results of the presented model may differ from other types of attacks. • Substantial computational resources are required for deep learning algorithms.	• Improve detection capabilities by combining traditional Rule-based IDS with ML algorithms. • Implement an anomaly detection system to identify malicious patterns. • Deploy lightweight ML models to reduce the computational cost and effectively run in resource-constrained IoT devices.
[12]	• The ability of ML detection and responsive capabilities has significantly improved. • Automation cybersecurity systems can be efficiently deployed through ML.	• ML is relaying on the data quality and quantity. • Vulnerabilities such as adversarial at- tacks were found in ML models.	Robust ML algorithms by using different datasets to gain more knowledge about cyber-attacks. Implement multiple defense systems to detect and respond against cyber-attacks.
[7]	• NID-Shield showed high accuracy levels and low false positive rates. These results were tested using UNSW-NB15 and NSL-KDD datasets. • CAPPER algorithm has improved the performance of intrusion detection.	• Only specific datasets were applicable for this system. • It also requires further investigation to measure the adaptability of real-time attack scenarios.	• Explore and deploy more datasets that include real-time traffic data from other environments. • Deploy data analysis tools to analyze network traffic in real-time such as Cassandra. • Design the system to operate in different nodes to handle overload network traffic.
[30]	• The key finding was that the random forest method • achieved over a 99% accuracy rate among other datasets.	• The model efficiency may differ across variety datasets.	• Adjust the detection thresholds to balance the sensitivity and specificity of operational environments.

		•The requirements of computing re- sources are limiting real-time integration. • Both false positives and negatives are still occurring.	• Combining anomaly detection with signature-based techniques will develop the detection capabilities.
[31]	• The successful combination between the GIWRF and Decision Tree model. • Outperformed model among others was the GIWRF-DT.	• The results depend on specific datasets and may not be applicable to other datasets. • The paper was focused only on binary classification.	• Develop approaches that are applicable for multiple classifications to be familiar with different types of intrusions. • To accurately identify relevant features for intrusion detection, it's important to utilize hybrid approaches that combine multiple selection methods.
[32]	• The proposed hybrid model has achieved a 99.9% accuracy rate for binary classification. • The model achieved 99.96% accuracy for multi-classification. • The hybrid model proved to be more effective than ML and DL algorithms.	• The paper has not addressed the issues of false positives and negatives in at- tack detection. • The paper does not discuss the scalability of the applied solution in real- world scenarios.	• It's recommended to design the IDS to be more modular and scalable to be able to adequately process data overload. • To improve the IDS efficiency, it's recommended to be evaluated under realistic scenarios.
[33]	• The proposed model achieved a range between 99.79% and 100% for detection accuracy rates. • It showed effectiveness compared to traditional IDS methods.	• The model dependent on specific datasets. Which may not be applicable for other types of threats. • Performance may be affected due to unaddressed L1 or L2 regularization of the model hyper parameters.	• To avoid overfitting, it's recommended to implement L1 and L2 regularization to improve hyperparameters of the DCNN model. • Expanding the training on various datasets will help to robust the model efficiency.
[2]	• The implementation of both multiple classifiers and feature selection has enhanced the IDS capabilities. • The proposed model has the best performance compared to other techniques.	• There are scalability limitations when used in complex IoT systems. • It is based on specific datasets that may minimize generalizability to realistic events.	• It's recommended to adapt a scalable distributed architecture to avoid scalability issues. • Implement processing tools to allow actual time data processing and anomaly detection. • Use a variety of datasets to train the model.
[34]	• The model has achieved a remarkable accuracy rate of 95.6%. • The architecture showed significant improvements in multiclass classification tasks.	• A lower rate of detection was observed for underrepresented classes due to an imbalance class in the datasets. • The model was trained on historical data, which may	• It's recommended to address the class imbalance by utilizing techniques such as oversampling to balance the attack types represented.

		not adapt to newly emerging threats. • Relying on specific datasets may limit the model's ability to be generalized to different environments.	• Retraining the model to be able to adapt with recent and evolving attack data. • The model should be tested with multiple datasets to evaluate its performance.
[35]	• The key findings are that the model has achieved a 99.9% accuracy rate. • Additionally, it detects various types of IoT attacks.	• The proposed model may not address all types of attacks adequately. • The model depends on specific datasets that may not be applicable to different environments.	• Ensure that the IDS is updated regularly and includes the latest attack data to maintain its effectiveness. • Use strategies such as data augmentation to simulate different attack events that might not be present in the selected dataset.

7. Discussion

The role of ML in enhancing the capabilities of IDS in detecting and responding is significantly effective. Anomaly detection is one of the ML features that is successfully used in this area. It enables the identification of data deviation by learning normal network behavior [12]. ML can detect any suspicious behavior that may contain intrusions. Addi- tionally, it allows a real-time analysis of enormous amounts of data. Accuracy of IDS can be ensured by minimizing the false positives and developing the alerts to be more operative [37]. ML can implement adaptive learning to empower the capability of IDS. By retraining the approaches with evolving threats, it will ensure an effective detection system. Due to the enormous amount of data that is being gathered from multiple sources, it develops the threat intelligence features. Robust prediction analytics will help organizations take proactive measurements against threats. It also requires having more accurate results when measuring threats [32].

7.1. ML Enhancements in IDS

ML can automate the responding system to directly act on identified threats. Addi- tionally, the system can learn from previous incidents to empower the response mechanism. By applying auto-ML that utilizes multiple types of ML approaches on datasets for training and tolerating actions based on the threats [28]. Additionally, it enables the monitoring of devices that are connected to the network. ML- based IDS will Look for and monitor any suspicious behavior within these devices [38]. It can also monitor the user behavior to detect insider risks. One of the significant features of enabling ML-based IDS is that it doesn't require pre-given signatures of known threats. Which considered an effective method against evolving and zero day-attacks. ML can deploy multiple ML approaches that specifically work as a detection system against various threats. According to [14], the significant approach for anomaly detection system that combines multiple ML algorithms. Such as deep neural network (DNN), one- class SVMs, and VAEs. These modes provide enhancements to an anomaly detection systems and reducing the false positive and negative rate. ML models have the ability to adapt to new evolving threats continuously. This will help IDS to keep a step ahead when potential vulnerability is encountered. This can be done by training ML models with training data and classifying it into supervised, unsupervised, and semi-supervised [39]. Analyzing smart cities, big data can be easily gained. By the significant role of ML algorithms that have the ability to handle complex and high transactions effectively. Additionally, integrating federated learning allows for training models across multiple devices without centralized data, enhancing privacy and scalability.

Once a security incident has occurred, ML can automatically generate reports. It will help to provide an accurate analysis of the situation and ensure that compliance is maintained. For instance, the R tool can be used to analyze incidents by using the time series feature [25]. Additionally, it will nuance the visualization of the data to be more understandable for users. Tools like Tableau [24] can provide visualization systems that help users understand patterns in network data flowing. ML algorithms can be integrated with other security tools. Integrating System Information and Event Management (SIEM) can

be done with ML to enhance the accuracy of threat detection systems [40]. Integrating ML with threat intelligence platforms such as the Malware Information Sharing Platform (MISP) will empower the detection capabilities efficiency [27]. By integrating ML into IDS, it will provide scalability, adaptability, and a detection system. Taking advantage of these features will help to provide a robust security framework for organizations. This framework will be proactive solutions and responsive to the continuously evolving threats.

7.2. Comparison with Other Review Papers

The aim of this study is to explore the connection between ML and the IDS. The primary focus is to analyze the security challenges related to utilizing traditional IDS. Com- paring to other studies that did not explore these security issues. Implementing a robust IDS with the integration of ML algorithms will address these problems. The study is exploring and investigating the possibility of utilizing new ML algorithms in IDS. Additionally, the purpose of this study is to highlight the advantages and drawbacks of adapting ML in IDS. By creating opportunities for future research and practical improvements. The study provides a comprehensive evaluation of relevant studies. Highlight the existing limitations and clarify the future works of integrating ML in IDS. In our paper, we investigate the different ML algorithms that have been used in this field. While other studies presented the approaches separately on different platforms. Our paper explored the role of ML in enhancing the IDS. Discussed the tools used for analyzing big data and how to mitigate these risks.

In this study [38], the paper concentrated on IoT networks without a specific urban area that limits the applicability in smart city real-world scenarios. Also, the paper may not address the full range of smart city environments. The paper focused on detection accuracy in a controlled environment. However, it didn't address the challenges related to real-time processing. Moreover, it utilizes a specific combination of DL and three-level algorithms. This may not include all possible approaches. The paper suggests extensions; however, it may not address scalability challenges in depth in smart city context.

In this study [2], the paper is focused on improving anomaly detection in IDS by using ML and ensemble techniques in IoT applications in smart cities. The paper relied on a limited dataset that could potentially lead to prejudiced results. The paper lacks the variety of the types of attacks analyzed and might not cover all real-world vulnerabilities. The paper primarily focused on batch processing and offline analysis. That leads to a lack of real-time processing capabilities, which impacts the systems usability in dynamic and rapidly changing environments. Additionally, the paper does not adequately address the scalability that makes it struggle with performance due to the data increment. Additionally, the paper focused on using a single technique that may miss critical aspects of IDS. It also lacks a comprehensive analysis of various attack vectors and defensive strategies.

Table 3. Comparison between This Study and Other Related Papers for Analyzing Big Data Using ML

Criteria	Our Paper	[38]	[2]
Data Diversity	√	×	×
Real-time Processing	√	×	×
Scalability	√	×	×
Comprehensive approach	√	√	×

8. Recommendations and Future Directions

8.1. Enhancing Organizations IDS with ML

ML is being deployed increasingly into IDS to the ability to detect and respond to threats. ML can train large datasets with normal and malicious network traffic. Thus, IDS can learn to identify abnormal patterns of attacks. This will allow us to identify zero-day attacks and new attack strategies in real time. Moreover, IDS accuracy can be improved using ML by reducing false positive and negative rates. Which results in more effective and efficient proactive measures. Following are some of the steps that organizations can take to improve their IDS performance using ML.

8.1.1. Model Training

Training ML models on real-time data is important to improve the detection capabil- ities of an IDS. By constantly revealing the model to the latest network traffic. This will allow the model to adapt to evolving threat scenarios. Also, the model will be able to identify the emerging attack strategies. This step will enable IDS to respond to new threats proactively and reduce

any potential security attacks. Real-time training helps the model to clearly understand the normal network patterns. By reducing the false positive and negative rates and improving detection accuracy.

8.1.2. Ensemble Learning Techniques

This technique provide a sufficient approach to improve IDS performance. This technique can improve detection accuracy, minimize false positives, and make the system robust. This can be achieved by combining multiple ML models. Applying these techniques has created robust performance due to the diversity of solo models. Ensemble techniques can handle complex and evolving threats effectively. Additionally, one of the common issues in ML models is overfitting, which can be reduced by ensemble techniques. It can improve the IDS to detect unseen threats.

8.1.3. Data Visualization and Incident Reporting Systems

These tools are considered important tools of an effective ML-based IDS. These tools provide a clear visualization of network traffic. This will help organizations identify anomalies and potential threats. Additionally, it will provide a deeper insight into the causes of attacks and help in decision-making. By enhancing incident report it will provide significant benefits for organizations. Tracking and analyzing security events, identifying threats, and posture the overall security. Integrating these technologies will improve the effectiveness of organizations IDS. Also, it will respond effectively to security threats and vulnerabilities.

8.1.4. Incident Response Automation

This method is important to evolve the effectiveness of ML-based IDS. This method- ology will automate threat identification, alert generation, and response actions. Orga- nizations can benefit from this method by reducing the response time and the impact of security threats. By automating response systems, it will fasten the ability to analyze alerts. Prioritize incidents based on the severity of the impact on the data. Also, it provides appropriate countermeasures based on the incident events. This will help the organizations systems to enhance the overall security operations.

8.2. ML Algorithms for Anomaly Detection

Anomaly detection is an important method in ML, it significantly aims to identify deviations from normal data. Multiple ML approaches are deployed to detect anomalies such as errors and suspicious activities. For instance, contrastive learning technique is a promised approach. It works by learning broad representations of normal data activities. This will force the model to learn features from the underlying structure of the data captured. It can be a robust tool for anomaly detection. This allows the model to adapt to the data normality [41]. Additionally, it's a self-supervised technique that doesn't require any explicit labels for normal and anomalous data. This is helpful when some domains labeled anomaly data is scarce.

Recent papers have considered Temporal Convolutional Networks (TCNs) as a promising approach. It is basically a deep learning architecture for modeling sequential data. Additionally, TCNs can be used for anomaly detection tasks. Due to its capability to efficiently capture patterns and temporal relationships in time series data [42]. Integrating TCNs in anomaly detection will identify deviations from normal data patterns. By training the network on historical data and pointing out instances that deviate from the learned patterns. TCNs provide multiple advantages, such as efficient computation and robust performance in capturing complex temporal connections [43]. Multimodal anomaly detection can extract information from multiple sources of data. Such as sensor data, text, or images, each has valuable information. This step will help to enhance the accuracy and efficiency of anomaly detection systems. Implementing multiple modalities helps to capture patterns and connections that might not appear in a single modality. Multimodal approaches are providing a clear sight of the underlying data. Additionally, it leads to more accurate and reliable detection results [44].

Future research could investigate the advantages of various methods to understand the full potential of ML algorithms in anomaly detection. Such as contrastive learning and TCNs, it will provide a powerful and strong anomaly detection system. Additionally, prioritize developing anomaly detection to continuously adapt to changing data. This will allow the model to be more efficient in detecting security threats. Future research should include more exploration of multimodal anomaly detection systems. This will improve the accuracy and efficiency of the anomaly detection system.

8. Conclusion

In conclusion, integrating ML into IDS has significantly enhanced cybersecurity in smart cities. As cities become more interconnected, advanced defenses are needed to combat the evolving cyber threats. This review highlights how ML techniques improve IDS capabilities effectively. Such as supervised, unsupervised, and reinforcement learning. By effectively detecting anomalies, reducing false positives, and activating timely alerts. However, there are still security challenges in spite of these advancements. Additionally, there is a need for robust configurations and algorithms to address evolving threats. Future research should be focused on improving these models. Moreover, enhancing real-time data processing is also imperative. By collaborating and sharing threat intelligence, a resilience security framework can be achieved. These frameworks will protect the critical infrastructure of smart cities. Additionally, it will ensure their security and reliability in an increasingly digital environment.

Corresponding author
Dr. M. M. Hafizur Rahman
mhrahman@kfu.edu.sa

Acknowledgements
NA.

Funding
No funding.

Contributions
R.A; M.M.H.R; Conceptualization, R.A; M.M.H.R; Investigation, R.A; M.M.H.R; Writing (Original Draft), R.A; M.M.H.R; Writing (Review and Editing) Supervision, M.M.H.R; H.A.A; Project Administration.

Ethics declarations
This article does not contain any studies with human participants or animals performed by any of the authors.

Consent for publication
Not applicable.

Competing interests
All authors declare no competing interests.

References

[1] Jan, M. A., He, X., Song, H., & Babar, M. (2021). Editorial: Machine learning and big data analytics for IoT-enabled smart cities. *Mobile Networks and Applications, 26*(1), 156–158. https://doi.org/10.1007/s11036-020-01702-4
[2] Bukhari, O., Agarwal, P., Koundal, D., & Zafar, S. (2023). Anomaly detection using ensemble techniques for boosting the security of intrusion detection system. *Procedia Computer Science, 218*, 1003–1013. https://doi.org/10.1016/j.procs.2023.01.080
[3] Musleh, D., Alotaibi, M., Alhaidari, F., Rahman, A., & Mohammad, R. M. (2023). Intrusion detection system using feature extraction with machine learning algorithms in IoT. *Journal of Sensor and Actuator Networks, 12*(2), 29. https://doi.org/10.3390/jsan12020029
[4] Abdallah, E. E., Eleisah, W., & Otoom, A. F. (2022). Intrusion detection systems using supervised machine learning techniques: A survey. *Procedia Computer Science, 201*, 205–212. https://doi.org/10.1016/j.procs.2022.03.029
[5] Zakaria, R., Abdelmajid, H., & Zitouni, D. (2022). Deep learning in medical imaging: A review. In *CRC Press eBooks* (pp. 131–144). https://doi.org/10.1201/9781003269793-15
[6] Vanin, P., Newe, T., Dhirani, L. L., O'Connell, E., O'Shea, D., Lee, B., & Rao, M. (2022). A study of network intrusion detection systems using artificial intelligence/machine learning. *Applied Sciences, 12*(22), 11752. https://doi.org/10.3390/app122211752
[7] N, T. R., & Gupta, R. (2021). Design and development of an efficient network intrusion detection system using machine learning techniques. *Wireless Communications and Mobile Computing, 2021,* Article 9974270. https://doi.org/10.1155/2021/9974270
[8] Aljanabi, M., Ismail, M. A., & Ali, A. H. (2021). Intrusion detection systems, issues, challenges, and needs. *International Journal of Computational Intelligence Systems, 14*(1), 560. https://doi.org/10.2991/ijcis.d.210105.001
[9] Mehmood, Y., Habiba, U., Shibli, M. A., & Masood, R. (2013). Intrusion detection system in cloud computing: Challenges and opportunities. *[Publication details missing]*.

[10] Celdrán, A. H., Sánchez, P. M. S., Castillo, M. A., Bovet, G., Pérez, G. M., & Stiller, B. (2022). Intelligent and behavioral-based detection of malware in IoT spectrum sensors. *International Journal of Information Security*, *22*(4), 541–561. https://doi.org/10.1007/s10207-022-00602-w
[11] Johnson, J., Jones, C. B., Chavez, A., & Hossain-McKenzie, S. (2023). SOAR4DER: Security orchestration, automation, and response for distributed energy resources. In *Distributed Energy Resources* (pp. 387–411). Springer. https://doi.org/10.1007/978-3-031-20360-2_16
[12] Sarker, I. H. (2022). Machine learning for intelligent data analysis and automation in cybersecurity: Current and future prospects. *Annals of Data Science, 10,* 1473–1498. https://doi.org/10.1007/s40745-022-00444-2
[13] Jogin, M., Manjunath, M., & others. (2018). Feature extraction using convolution neural networks (CNN) and deep learning. In *IEEE Conference Publication*. IEEE.
[14] Akinola, O., Akinola, A., Ifeanyi, I., Adewole, O., Sulaimon, B., & Oyekan, B. (2024). Artificial intelligence and machine learning techniques for anomaly detection and threat mitigation in cloud-connected medical devices. *International Journal of Scientific Research and Modern Technology*, *3*(3), 1–13. https://doi.org/10.38124/ijsrmt.v3i3.26
[15] Lesouple, J., Baudoin, C., Spigai, M., & Tourneret, J. Y. (2021). Generalized isolation forest for anomaly detection. *Pattern Recognition Letters, 149,* 109–119. https://doi.org/10.1016/j.patrec.2021.05.022
[16] Togbe, M. U., Barry, M., Boly, A., Chabchoub, Y., Chiky, R., Montiel, J., & Tran, V. T. (2020). Anomaly detection for data streams based on isolation forest using Scikit-Multiflow. In *Advances in Intelligent Systems and Computing* (pp. 15–30). Springer. https://doi.org/10.1007/978-3-030-58811-3_2
[17] Alavizadeh, H., Alavizadeh, H., & Jang-Jaccard, J. (2022). Deep Q-learning based reinforcement learning approach for network intrusion detection. *Computers, 11*(3), 41. https://doi.org/10.3390/computers11030041
[18] Gronauer, S., & Diepold, K. (2021). Multi-agent deep reinforcement learning: A survey. *Artificial Intelligence Review, 55,* 895–943. https://doi.org/10.1007/s10462-021-09996-w
[19] Wang, Y., & Zou, S. (2022). Policy gradient method for robust reinforcement learning. *[Publication details missing]*.
[20] Ali, W. A., N, M. K., Aljunid, M., Bendechache, M., & Sandhya, P. (2020). Review of current machine learning approaches for anomaly detection in network traffic. *Journal of Telecommunications and the Digital Economy, 8*(4), 64–95. https://doi.org/10.18080/jtde.v8n4.307
[21] Duong, H. T., Le, V. T., & Hoang, V. T. (2023). Deep learning-based anomaly detection in video surveillance: A survey. *Sensors, 23*(11), 5024. https://doi.org/10.3390/s23115024
[22] Ullah, A., Anwar, S. M., Li, J., Nadeem, L., Mahmood, T., Rehman, A., & Saba, T. (2023). Smart cities: The role of Internet of Things and machine learning in realizing a data-centric smart environment. *Complex Intelligent Systems, 10*(3), 1607–1637. https://doi.org/10.1007/s40747-023-01175-4
[23] Amović, M., Govedarica, M., Radulović, A., & Janković, I. (2021). Big data in smart city: Management challenges. *Applied Sciences, 11*(10), 4557. https://doi.org/10.3390/app11104557
[24] Cesario, E. (2023). Big data analytics and smart cities: Applications, challenges, and opportunities. *Frontiers in Big Data, 6,* 1149402. https://doi.org/10.3389/fdata.2023.1149402
[25] Nuaimi, E. A., Neyadi, H. A., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications, 6*(1), 25. https://doi.org/10.1186/s13174-015-0041-5
[26] Brahim, M. B., Drira, W., Filali, F., & Hamdi, N. (2016). Spatial data extension for Cassandra NoSQL database. *Journal of Big Data, 3*(1), 11. https://doi.org/10.1186/s40537-016-0045-4
[27] Kasongo, S. M. (2023). A deep learning technique for intrusion detection system using a recurrent neural networks based framework. *Computer Communications, 199,* 113–125. https://doi.org/10.1016/j.comcom.2022.12.010
[28] Xu, H., Sun, Z., Cao, Y., & Bilal, H. (2023). A data-driven approach for intrusion and anomaly detection using automated machine learning for the Internet of Things. *Soft Computing, 27,* 14469–14481. https://doi.org/10.1007/s00500-023-09037-4
[29] Campos, E. M., Saura, P. F., González-Vidal, A., Hernández-Ramos, J. L., Bernabé, J. B., Baldini, G., & Skarmeta, A. (2022). Evaluating federated learning for intrusion detection in Internet of Things: Review and challenges. *Computer Networks, 203,* 108661. https://doi.org/10.1016/j.comnet.2021.108661
[30] Hossain, M. A., & Islam, M. S. (2023). Ensuring network security with a robust intrusion detection system using ensemble-based machine learning. *Array, 19,* 100306. https://doi.org/10.1016/j.array.2023.100306
[31] Disha, R. A., & Waheed, S. (2022). Performance analysis of machine learning models for intrusion detection system using Gini impurity-based weighted random forest (GIWRF) feature selection technique. *Cybersecurity, 5,* 1. https://doi.org/10.1186/s42400-021-00103-8
[32] Yaras, S., & Dener, M. (2024). IoT-based intrusion detection system using new hybrid deep learning algorithm. *Electronics, 13*(6), 1053. https://doi.org/10.3390/electronics13061053
[33] Hnamte, V., & Hussain, J. (2023). Dependable intrusion detection system using deep convolutional neural network: A novel framework and performance evaluation approach. *Telematics and Informatics Reports, 11,* 100077. https://doi.org/10.1016/j.teler.2023.100077
[34] Ashiku, L., & Dagli, C. (2021). Network intrusion detection system using deep learning. *Procedia Computer Science, 185,* 239–247. https://doi.org/10.1016/j.procs.2021.05.025
[35] Kayode Saheed, Y., Idris Abiodun, A., Misra, S., Kristiansen Holone, M., & Colomo-Palacios, R. (2022). A machine learning-based intrusion detection for detecting Internet of Things network attacks. *Alexandria Engineering Journal, 61*(12), 9395–9409. https://doi.org/10.1016/j.aej.2022.02.063

[36] Awajan, A. (2023). A novel deep learning-based intrusion detection system for IoT networks. *Computers, 12*(2), 34. https://doi.org/10.3390/computers12020034
[37] Brahim, M. B., Drira, W., Filali, F., & Hamdi, N. (2016). Spatial data extension for Cassandra NoSQL database. *Journal of Big Data, 3*(1), 11. https://doi.org/10.1186/s40537-016-0045-4
[38] Alosaimi, S., & Almutairi, S. M. (2023). An intrusion detection system using BoT-IoT. *Applied Sciences, 13*(9), 5427. https://doi.org/10.3390/app13095427
[39] Logeswari, G., Bose, S., & Thangasamy, A. (2023). An intrusion detection system for SDN using machine learning. *Intelligent Automation & Soft Computing, 35*(1), 867–880. https://doi.org/10.32604/iasc.2023.026769
[40] Sasi, T., Lashkari, A. H., Lu, R., Xiong, P., & Iqbal, S. (2024). A comprehensive survey on IoT attacks: Taxonomy, detection mechanisms and challenges. *Journal of Information and Intelligence, 2*(4), 455–513. https://doi.org/10.1016/j.jiixd.2023.12.001
[41] Sabiri, B., Khtira, A., Asri, B. E., & Rhanoui, M. (2024). Investigating contrastive pair learning's frontiers in supervised, semisupervised, and self-supervised learning. *Journal of Imaging, 10*(8), 196. https://doi.org/10.3390/jimaging10080196
[42] Qiu, L., Jin, L., & Chai, L. (2023). Network traffic prediction based on spatio-temporal graph convolutional network. In *Proceedings of the 2023 42nd Chinese Control Conference (CCC)* (pp. 8426–8431). IEEE. https://doi.org/10.23919/CCC58697.2023.10239918
[43] Park, J., Park, Y., & Kim, C. I. (2022). TCAE: Temporal convolutional autoencoders for time series anomaly detection. In *Proceedings of the 2022 Thirteenth International Conference on Ubiquitous and Future Networks (ICUFN)* (pp. 421–426). IEEE. https://doi.org/10.1109/ICUFN55119.2022.9829692
[44] Zhao, Z., & Chen, M. (2024). Time series anomaly detection and prediction model integrating multimodal data. In *Proceedings of the 2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)* (pp. 1–5). IEEE. https://doi.org/10.1109/IACIS61494.2024.10721738