Jurnal Komputer, Informasi dan Teknologi Vol: 5, No 1, 2025, Page: 1-19

Securing Smart Cities - The Impact of Machine Learning Technology

Hazem Salim Abdullah[1]*, Salim Abdullah Saleh[2]

[1]Directorate of Municipalities Nineveh Governorate, Mosul, Iraq
[2]Retired Assist. Prof. Dr., Tikrit University, Eng. College, Mosul, IRAQ

DOI: https://doi.org/10.53697/jkomitek.v5i1.2748
*Correspondence: Hazem Salim Abdullah
Email: hazemabdullah956@gmail.com

Received: 14-04-2025
Accepted: 22-05-2025
Published: 30-06-2025

Abstract: Environmental concerns have led to a growing focus on sustainability in computing systems, from small devices to big data centers. At the same time, technologies like resource-intensive encryption and sophisticated intrusion detection systems are necessary to provide cybersecurity in these linked networks. Artificial intelligence (AI) is changing how the world works, and innovative products are the end consequence. AI and the Internet of Things (IoTs) enable several brilliant advances that make up a smart city. The concept highlights a smart city's simplicity and comfort, but its growth is hampered by several security issues that are being brought up. An intrusion detection system (IDS) monitors network traffic and notifies users of abnormalities. A method based on machine learning, IDS makes decisions regarding the legibility of data packets, detects network threats intelligently, and notifies the user. Researchers have used a variety of machine-learning techniques to increase the detection accuracy of IDS. A comparison of several machine learning algorithms trained over the NSL-KDD dataset. XG Boost, Random forest K-Nearest Neighbors KNN and Logistic Regression are simulated, which generates excellent performance accuracies (99.92%, 99.86%, 99.57%, and 88.05%). It has been observed that the LR model has the lowest accuracy among the different models, while the XGB model has the highest accuracy.

Keywords: Machine Learning, Intrusion Detection System, IoT network, Smart Cities

Introduction

Technological development has changed the dynamics of the world. IoT and wireless communication networks are driving the automation of infrastructure in all sectors. Smart cities depend on wireless communication; without infrastructure, several cyberattacks are possible. Therefore, weaknesses in smart cities have to be addressed with suitable remedies. Smart cities cover many applications, including e-government, innovative housing, intelligent transportation, telemedicine, smart grid, UAV surveillance, energy, and many more (Cimen et al., 2020; Panagiotis et al., 2021). Due to the ever-increasing number of cyber-attacks, many academics worldwide focus on data network security. Intrusion detection is a system that identifies fraudulent data packets with ease. The optimal IDS algorithm balances high accuracy with false negative and false positive metrics. Also, the primary

objective of IDS is to identify potential cyber-attacks. However, intrusion detection systems are based on legitimate and unlawful data packets.

Furthermore, smart cities require secure communication routes, and IDS plays a crucial role (Liao et al., 2021; Rincy & Gupta, 2021). Figure 1 depicts the concept of smart cities includes intelligent houses, healthcare facilities, and vehicles, as well as connectivity inside a smart city. Unmanned Aerial Vehicles can be incorporated into smart cities to enhance connectivity. Secure communication networks aim to minimize end-to-end latency. An intruder can employ false data injection attacks to disrupt communication during remote surgery on a high-profile individual. Diverse technologies, including Markov chains, machine learning, deep learning, ant colony optimization, and Poisson distribution, are employed to enhance signatures.

Literature Survey

This section discusses IoT technology techniques and intrusion detection systems that use machine learning and deep learning algorithms to enhance security. The rapid spread of IoT environments equipped with sensors, actuators, and CPUs necessitates the use of intrusion detection systems in both traditional and modern industrial automation. The Industrial Internet of Things is a complex system; each error or violation can cause significant damage. Therefore, fast and reliable cyber-attack identification is crucial for successful network response. An intrusion detection system protects network data by detecting network breaches quickly.

In 2017, Zhou & Paffenroth proposed a random forest algorithm technique for selecting important features, achieving an accuracy of 0.9933. Using support vector machine group learning (SVM) and logarithmic marginal density ratios transformation (LMDRT), other authors (Gu et al., 2019) developed an intrusion detection model in 2019 to enhance the quality of the data. According to CICIDS2017 data, the accuracy is 0.9364, the DMA is 0.9756, and the far is 0.2028. Also, Wang et al. proposed a multi-factor reinforcement learning model for intrusion detection systems, in 2020 to focusing on reinforcement learning.

To safeguard Industry 4.0 against cybersecurity mishaps, assaults on automated framework development, and computer hacking of SCADA systems, Chen and colleagues created NIDS, a CNN-based data collecting and surveillance management system, in 2020. Yuan et al. presented the DeepFed method in 2021 to detect and thwart cyber threats that divert IoT devices.

These approaches, however, are inadequate for handling global data that is big, complicated, and multivariate. Working with IoT data often requires a significant learning program, and accuracy should be improved. Gee et al. investigated IDS technology for IoT systems using levels and three thick layers in 2021. BoT-IoT was employed for multi-class identification to differentiate between regular and semi-malicious private invasions, achieving an accuracy of 0.9979. Ullah and Mahmoud suggested a convolutional neural network-based deep learning model, in 2021 for an intrusion detection system, achieving a low detection rate of around 0.997 for binary and multicast classifications. Al-Kasassbeh et al. showed that the LightGBM technique surpassed the DL approach. Guezzaz et al.

introduced the NIDS model utilizing decision trees. They conducted a comparative analysis with analogous datasets, including the NSL-KDD and CICIDS2017, as a valuable foundation for evaluating and enhancing data quality. The model's overall accuracy on the NSL-KDD and CICIDS2017 datasets was 0.9942 and 0.9880, respectively.

Guezzaz et al. developed a novel detection technique using a multilayer classifier (MLP) to collect data packets. While traditional machine learning methods focus on feature selection, more recent studies emphasize reinforcement learning to protect IoT-based smart grids from botnet attacks. Ashraf et al. presented the robot-IDS intrusion detection system, a novel botnet detection system based on experience learning. Using experience learning based techniques, such as the symmetry model and the beta-mixture model (BMM), the system builds models of the expected behaviour of IoT networks. Any departure from normal behaviour is considered an abnormal occurrence. Three reference datasets derived from native IoT networks were utilized to evaluate the iotbot-IDS intrusion detection system. The assessment findings show that, with an overall detection rate of 0.992, the robotIDS intrusion detection system effectively detects different types of botnets.

In 2021& 2022, Kasongo & Mohy-Eddine et al. introduced an effective anomaly detection technique utilizing group learning to enhance the security of Edge Computing IIoT. They employed RF, LR, NB, DT, pet, and GB technologies for characteristic geometry and RF for the fitness function. Their model attained an average accuracy of 0.8761 and a sub-AUC area of 0.98.

Guezzaz et al. used machine learning techniques, in 2021 to create an edge-based IoT security anomaly detection system. This method identified abnormalities and overuse using fundamental component analysis (PCA) and the k-nearest neighbour (KNN) technique. The model achieved an accuracy of 0.98, precision of 0.97, and false acceptance rate (FAR) of 0.029 on the BoT-IoT dataset and an efficiency of 0.99, precision of 0.98, and FAR of 0.027 on the NSL KDD dataset.

To detect injection assaults in smart cities, Jaber et al. & Guezzaz et al. utilized a variety of feature selection approaches, including machine learning classifiers like SVM, RF, and DT, as well as repeated feature deletion and static removal. The DT Model can detect injection assaults with 0.99 accuracy with just eight features gathered using the particular feature selection method. Additionally, IDS-SIoEL, an intrusion detection system for IoT-based intelligent settings based on group learning and AdaBoost, was proposed by Hazman et al., in 2023. The model was assessed using the NSL-KDD, bot-IoT, and IoT-23 datasets, incorporating Boruta, mutual information, and correlation.

Using a KNN classifier and feature selection, Mohy-Eddine et al. created a network intrusion detection (NID) model for Internet of Things environments. By employing PCA, single-variable statistical testing, and the general feature selection technique to enhance data quality and choose the top ten performance characteristics, this model improved ID accuracy and detection rate (DR). Douiba et al. used supervised and deep learning to develop a novel intruder detection system for IoT-based intelligent settings. The system uses MLP, SVM and KNN for classification, IG and GA for feature selection, and Stack AE for deep extraction. The proposed model metrics were validated using the BoT-IoT dataset

compared to previous anomaly detection methods. The performance metrics ACC, recall, and precision were used.

IoT-Network Intrusion Detection System

Due to their generally inadequate security, IoT networks are susceptible to various attacks on the vital data they transport. When an attack is suspected, an IDS is an alarm that beeps (Ayub et al., 2023). IDSs are more prevalent because they continuously monitor network traffic, ensuring that no unconfirmed packets go unchecked. Researchers are currently more interested in leveraging machine learning and deep learning algorithms to improve intrusion detection in IoT networks (Hodo et al., 2016; Roopak et al., 2020). IoT increases data exposure to cyber threats due to its direct relationship with potential attack points, weak protection, and sensitive data transmission over unsecured networks. Strong security standards like encryption, regular updates, and strong passwords should be adopted to reduce risks. Therefore, it is crucial to defend the infrastructure of smart cities from potential threats. Even on sensor nodes in IoT networks, detection has become possible due to advancements in IoT security (Jan et al., 2020). In IoT environments, detecting intrusions in sensor nodes with limited resources is a significant challenge. Recent advancements like edge computing and lightweight algorithms have made intrusion detection systems (IDS) more effective. These systems use lightweight, resource-efficient algorithms like DT and KNN to detect attack trends and encryption methods like AES-128 to protect data while using fewer resources. Edge computing nodes send computations from the cloud to adjacent devices, such as gateways or sensors, for pretrained intrusion detection models, distributed processing, and pre-filtering. Combining these methods improves security in complex and dispersed IoT systems by balancing performance and resources. The three popular intrusion detection systems—signature based, anomaly-based, and hybrid—will be covered in more detail in Section 6. A comparison between their functions is summarized in Figure 1.

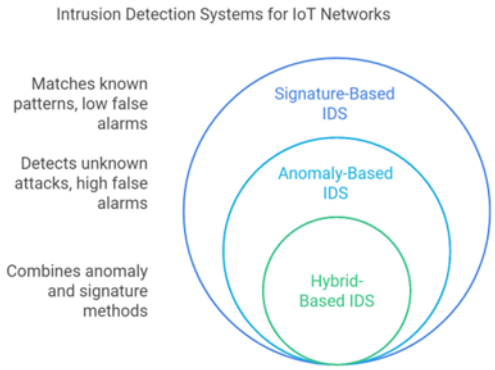

Figure 1. A comparison between functions of IDS types.

IoT-Network Real-Time Implementation in Smart Cities

Real-time IoT implementation faces challenges like latency, security, and scalability. However, solutions like edge computing, lightweight algorithms, and AI techniques can overcome these issues, enabling reliable and efficient performance of real-time applications. Protecting these networks is crucial for their effectiveness. As smart cities collaborate, IoT networks are central to automation, faster communication, and applications in intelligent

monitoring, traffic management, emergency response, healthcare, homes, grids, and agriculture (Zhou & Paffenroth, 2017). Practical solutions include developing 5G network infrastructure, using artificial intelligence applications for early detection, improving resource management with energy-efficient processors and DVFS technologies, and enhancing cloud and edge computing cooperation to reduce latency and increase efficiency (Gaur et al., 2015). The following sections of this chapter will provide a complete description of each of the Internet of Things application areas in smart cities, which are summarized by the author in Figure 2:

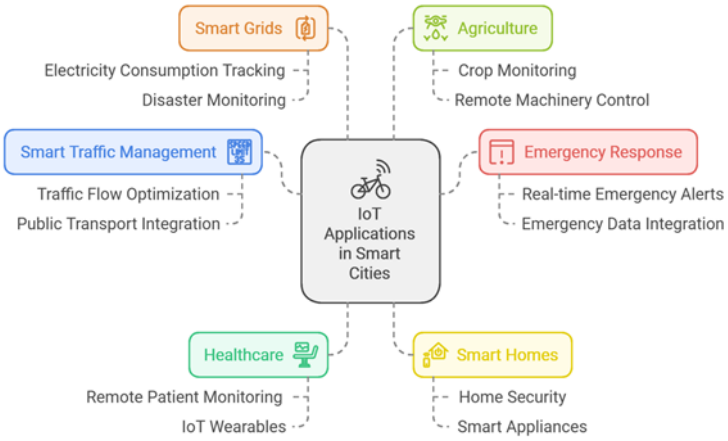

Figure 2. Domains of IoT Applications in Smart Cities and Their Primary Objectives

Smart Traffic Management.

Traffic flow, vehicle speeds, and density may all be tracked in real-time by IoT sensors on highways and crossings. This information may improve traffic flow and alleviate congestion, especially in smart cities during peak traffic periods, by optimizing traffic signals and routing (Fiore et al., 2019).

IoT can enhance smart traffic management by integrating traffic data with public transport systems, enabling efficient bus and train routing and scheduling. Real-time traffic data can also be integrated into journey-planning apps, enabling informed trip planning.

Emergency Reaction

Smart cities are using IoT networks for real-time emergency reaction by placing sensors in public buildings, streets, and parks to identify possible crises. These sensors may notify authorities when they detect smoke, fire, or very high or low temperatures. IoT networks also gather real-time data about emergency services vehicles, improving response efficiency and resource allocation (Mohammad et al., 2019).

By integrating emergency data with public alarm systems, IoT networks allow authorities to notify the public of possible dangers promptly and offer reaction instructions. By increasing the speed and effectiveness of emergency reaction activities in smart cities, this integration may result in safer smart city conditions and faster responses (Shah et al., 2019).

Healthcare IoT

IoTs have much to offer the healthcare industry, improving Medicare facilities even in remote locations for relatively little money. Real-time monitoring of patient's vital signs with smart wearables warns medical personnel of possible hazards. This enables medical personnel to keep an eye on patients without having to visit them in person or stay in the hospital. Thanks to IoT networks, results may be thoroughly examined (Pradhan et al., 2021).

A Smart Home powered by IoT

IoT has dramatically improved home automation since bright versions of equipment such as air conditioners, refrigerators, washing machines, fans, and door locks are now available. Vehicles have also been moved to IoT networks, which enable remote access. IoT sensors, such as gas fire and smoke sensors, keep an eye on houses and guard against possible dangers like fire and burglary, protecting occupants (Asadullah & Raza, 2016).

Smart Grid with IoT

A smart grid tracks usage at every system point and delivers power to consumers via two-way digital communication. IoT goals improve efficiency, accuracy, and operation duration compared to manual methods by providing disaster and operation monitoring on high-voltage transmission lines (Ou et al., 2012).

IoT for Agriculture

IoT is transforming agriculture by increasing production through smart farming and monitoring crops and animals remotely through tools and sensors connected to IoT networks. This technology also allows for remote control of machinery and the deployment of UAVs in modern farming and agriculture (Kim et al., 2020).

Smart City Cyberthreats

IoT networks are susceptible to cyber-attacks, posing significant challenges in mitigating such risks inside a smart city. DOS, DDOS, Sybil attacks, SQL injection, and malware assaults are prevalent in IoT contexts; thus, smart cities are vulnerable to these threats. If inadequately protected, an insecure sensor node network may lead to system failures or service interruptions. Such technological flaws can halt this development entirely. Fortunately, no one is powerless against these risks, as several diverse solutions are available for appropriate application (Chohan et al., 2023).

Smart city denial of service (DoS) attacks

A denial of service (DoS) attack is a fundamental attack that can render a target system inoperable or inaccessible, even for authorized users, due to the hacker's large amount of data packets. This attack aims to disrupt victim services, similar to a 2015 attack on a Smart Grid in Ukraine. Monitoring network traffic in smart cities is crucial to prevent

system outages and ensure safety (Sikder et al., 2021). Figure 3 illustrates the DoS attack technique.

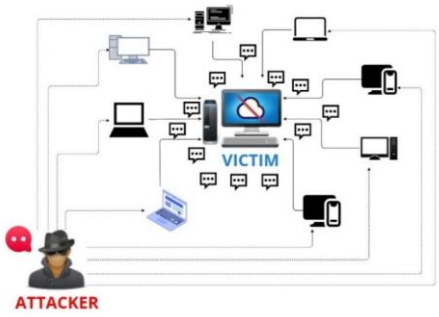

Figure 3. DoS attack mechanism

Smart Cities and Distributed Denial of Service Attacks (DDoS)

A distributed denial of service (DDoS) attack is a type of DoS attack that targets a single or multiple victims with infected computers or botnets operating across channels. This clever attack exhausts the server's or network infrastructure's resources, leaving the victim unhappy. Disrupting all communications can have adverse effects (Chohan et al., 2023). Figure 4 illustrates the attacker, handler, botnet, and victim in a DDoS assault on a smart city.

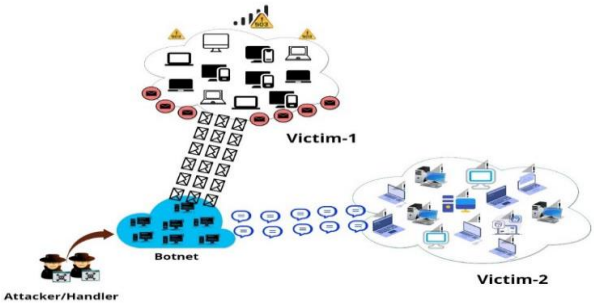

Figure 4. DDoS attacks on smart cities

Sybil Attack on Smart Cities

Sybil attacks involve hackers assuming multiple identities simultaneously, compromising system effectiveness. They can lead to privacy invasions, false reports, and spam in smart cities. They use tactics like malware, social engineering, and phishing and promote machine-learning techniques in their attack patterns. Figure 5 illustrates a Sybil Attack in an IoT environment (Chohan et al., 2023).

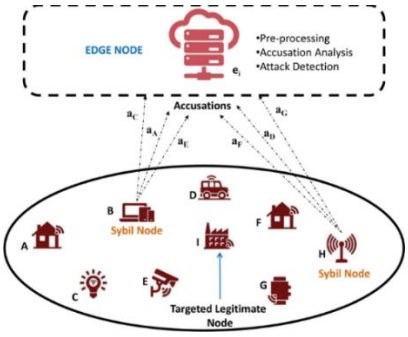

Figure 5. Sybil Attack in an IoT envirnment

SQL Injection Attack on Smart Cities

SQL injection is a well-known method for dealing with sensitive data. This incursion can damage SQL databases and read and erase data. In a smart city, all of the private information from different endpoints and sensor nodes of smart appliances may be in danger. Therefore, in order to preserve users' privacy, databases in smart cities need to be extremely secure (Sikder et al., 2021). Figure 6 illustrates an example of an SQL Injection Attack.

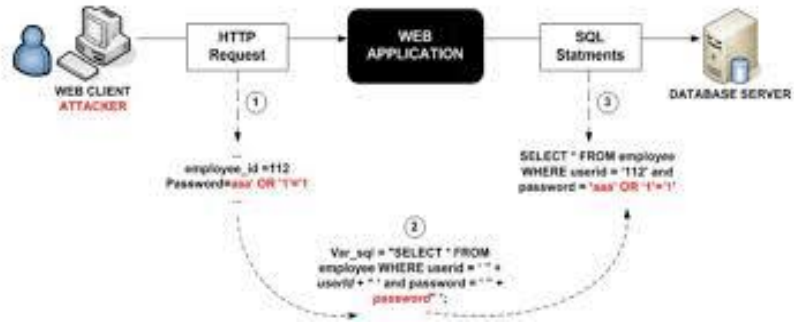

Figure 6. Example of a SQL Injection Attack

Malware Attack on Smart Cities

With various intrusion and threat types and classifications, malware is one of the most prominent groups of danger. Malware includes well-known types, including worms, Trojans, and ransomware. It causes data loss by infecting the victim with various viruses. Smart cities are the primary source of human comfort; however, all client data may be at risk of destruction (Garcia et al., 2009). The author summarized the most known types of malware in Figure 7.

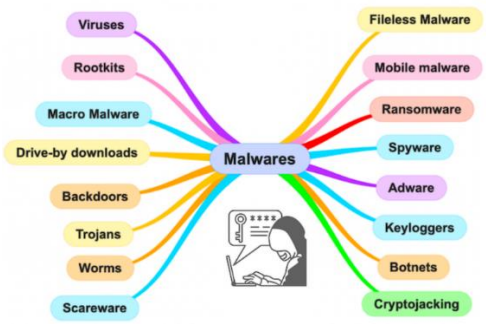

Figure 7. The Most Known types of Malware

Machine Learning-Based Intrusion Detection System

Computers may learn from data using a wide range of techniques known as machine learning. Machine learning is characterized by a set of essential characteristics, including:

- Its algorithm, which performs better with structured and ordered data, may be effectively built with a small to medium quantity of data.
- It breaks down the issue into minor matters and resolves each separately.
- Its algorithms, like decision trees and linear regression, have a fundamental structure.
- A central processing unit (CPU) is capable of running it.
- More human involvement is needed to choose and analyze characteristics in its algorithms to select the correct input.
- It takes less time during training but slows down during testing.
- Some machine learning algorithms remain the quickest despite this.

Therefore, Machine learning is becoming a crucial tool for identifying and reducing cyber threats in smart cities. Three main categories of machine learning techniques include anomaly-based, signature-based, and hybrid systems (Chohan et al., 2023). A hybrid system combines the strengths of both approaches for greater accuracy and efficiency. Anomaly-based detection employs system intelligence that has been educated using various methods, whereas signature-based detection compares network traffic to pre-existing attack patterns or signatures. Researchers have created multiple IDSs using different methodologies, algorithms, and target systems, comparing their accuracy and precision with other algorithms. Figure 8 illustrates a Machine-learning -based-intrusion-detection-system.

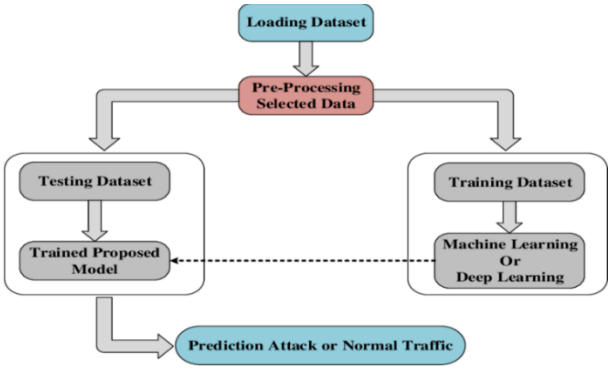

Figure 8. Machine-learning -based-intrusion-detection-system

We will provide a brief overview of the relevant IDS algorithms in this part, as follows:

XGBoost algorithm

XGBoost is a decision tree-based ensemble technique for regression and classification that improves the gradient's boost by fixing faulty estimators and gradually adding predictors. It considers past forecast outcomes and recent projections to reduce loss. XGBoost uses automated regularization to improve models' generalization performance, offering a more regularized version of Gradient Boosting. It enables rapid training and parallel processing across nodes, reducing training loss and regularization in its iterations (Hazem, 2023).

Random Forest algorithm

Random Forests are ensemble learning techniques used in regression and classification. During the preparation phase, several decision trees are constructed, and the final classification is determined by the majority vote of the classes generated by each tree. Furthermore, by calculating an average to establish a standard harmony between the two boundaries, they seek to mitigate the issues of significant variability and high susceptibility (Pai & Adesh, 2021).

K-NN algorithm

K-nearest neighbors evaluate newly acquired data according to how similar it is to previously collected data for regression and classification challenges. Some refer to the Knearest neighbor strategy as a lazy learner since the learning phase does not occur directly from the training set (Hazem, 2023). However, the approach retains the acquired data and uses matching criteria to classify the new incoming data. All that be done through successive steps of the algorithm where the number of points that are K neighbors is found in the first step, then the area in the neighborhood of each k-neighbor in the second step, Select K adjacent residents in the third step, and the fourth step Tally up all the data points for each category among the k-neighbors, and finally the class with the largest neighbor total is the one that comes out of the forecast.

Logistic Regression algorithm

Logistic Regression (LR) is a machine learning approach used for binary classification tasks and multiclass classification when using one-vs-rest approaches. It applies a linear model to the sigmoid function or its variants, squashing the output between 0 and 1 and determining a class's probability by an output closer to 1, which can then be mapped to two or more discrete classes (Pai & Adesh, 2021).

NSL-KDD Dataset

The KDD99 dataset was developed in 1999 and is one of the most popular study datasets in cyber security. Researchers have identified many problems that need to be addressed after carefully examining KDD99, including duplication and excessive items in the training and testing datasets. These issues present challenges when working with a complete dataset in studies.

A more recent version, NSL-KDD, was put forward to address the issues above (Abdullah, 2024). Since 2009, NSL-KDD has been recognized as the primary dataset for research on cyber security. The NSL-KDD dataset comprises two subsets: KDDTrain+, which includes 80% or 125,973 records, and KDDTest+, which shall consist of 20% or 22,544 records. Each record's 41 properties are divided into four categories: fundamental features, connection-based traffic features, time-based traffic features, and content features. Each record has twenty-one projected label classes assigned to it, representing both attack and regular recordings. In cyber security, every piece of information is regarded as a session, which is a connection between two hosts inside a network. The probability distribution of

KDDTrain+ and KDDTest+ differs. To evaluate the ability of models to generalize to the real world, NSLKDD calls for evaluating models on datasets that include attack types present in the test data but absent from the training data. This technique ensures that models can accurately identify ongoing attacks and adapt to emerging cyber threats. This example illustrates the challenges intrusion detection systems (IDS) face in the real world, where new attack variants are constantly being discovered. The test dataset contains 14 distinct attack types not included in the training dataset, while the training dataset contains 24 different attack types. This method tests the classifier's capacity to identify unknown assaults, guaranteeing that these models are accurate in identifying known attacks and flexible enough to adjust to emerging cyber threats. By introducing a novel technique, NSL-KDD enhances the KDD99 dataset. The NSL-KDD suite significantly improved KDD99 by addressing redundancy, bias, and inaccurate evaluation. It is a superior tool for intrusion detection systems development and contributes to research on more efficient cyber threat models. A detailed summary of the NSL-KDD record information is given in Table 1.

Table 1. NSL-KDD record Details

	All Records	Normal	DOS	Probe	R2L	U2R
KDDTrain+	125,973	67343	45927	11656	995	52
KDDTest+	22,544	9711	7458	2421	2754	200

Evaluation Metrics

- Accuracy in percentage terms is the number of data points that were accurately predicted out of all the data points. Equation 1 provides the accuracy calculation.

  Accuracy = (TP + TN) / (TP + FP + TN + FN) (1).

  The numbers of true positive instances (TP), true negative cases (TN), false positive cases (FP), and false negative cases (FN) are represented.
- Recall, also called sensitivity, is the percentage of positive class instances that are accurately anticipated to be positive. Equation 2 illustrates this formula.

  Recall = TP/(TP + FN) (2).
- Precision, whose formula is in Equation 3, calculates the probability that a positive prediction will come true. The optimal degree of accuracy is 1.

  Precision = TP / (TP + FP) (3).
- F1- Score is an indicator of the test's accuracy. It computes the score by considering the test's precision and recall. Equation 4 illustrates this formula.

  F1-Score = 2TP / (2TP + FP + FN) (4).
- The Specificity or False Positive Rate (FPR) is used to identify common assaults by dividing the number of false positive instances by the total number of negatives, with the optimal rate being 0.0 (see Equation 5 and Figure 9).

  Specificity or FPR = FP / (FP + TN) (5).

  Meanwhile, sensitivity (Recall) or true positive rate (TPR) is the percentage of positive cases the model properly classifies. The ideal value for it is 1.0.

  Sensitivity or TPR = TP / (TP + FN) (6).

  A point at (0, 1) can represent a perfect model with a TPR of 1.0 and FPR of 0.0 at a certain threshold.

Implementation of proposed ML models

The Python 3.7 language is used to create and implement the machine learning algorithms XGboost (XGB), Random Forest (RF), K-Nearest Neighbor (KNN), and Logistic Regression (LR). The system used for analysis is an i5 CPU, Windows 11 operating system, and 16 GB of RAM. The NSL-DKK dataset is divided into an 80:20 ratio for training and testing, ensuring a balance between bias and variance. The 20% test set provides a sufficient sample for generalization performance, while the 80% training set efficiently trains each model. Equations (1)– (6) are used to evaluate each model.

Methodology

This study employed a comparative analysis of machine learning algorithms to enhance intrusion detection in IoT-based smart city environments. Four widely used models—XGBoost (XGB), Random Forest (RF), K-Nearest Neighbors (KNN), and Logistic Regression (LR)—were implemented using Python 3.7. The NSL-KDD dataset, a refined version of KDD99 designed for cybersecurity research, served as the benchmark for training and evaluating these models. The dataset was split into 80% for training and 20% for testing to ensure generalization and reduce bias.

Each model was trained to classify network traffic as either normal or malicious based on 41 input features, categorized into basic, traffic-based, time-based, and content-based types. Model performance was evaluated using standard classification metrics: accuracy, precision, recall, F1-score, true positive rate (TPR), and false positive rate (FPR). The analysis was conducted on a system with an Intel i5 CPU, 16GB RAM, and Windows 11 OS.

By testing the models on attack types unseen in training data, the experiment assessed the algorithms' ability to detect zero-day threats. The models' predictive capabilities were compared, with XGB achieving the highest accuracy and LR the lowest. This methodology demonstrates the feasibility of using supervised machine learning to identify cyber threats in smart city IoT networks effectively.

Results and Discussion

- Tables 2 and 3 display the classification reports and different performance accuracies obtained from the confusion matrices shown in Figure 9. Different ML models' performances, shown in Figure 10, are measured after they have been trained using the training dataset, validated using the validation datasets, and then tested using the test dataset. Table 2 displays the various performance accuracies (99.92%, 99.86%, 99.57%, and 88.05%). It has been observed that the LR model has the lowest accuracy among the different models, while the XGB model has the highest accuracy.
- The XGB model outperforms RF, KNN, and LR models for intrusion detection or classification tasks on large, high-dimensional datasets due to its versatility, ability to manage complex linkages and nonlinear feature interactions, and iterative prediction improvement capabilities. It also has high accuracy, scalability to parallel processors with optimized hyperparameters, and the ability to handle outliers or missing data.

- In contrast, we observe that the FPR scale values approach 0 for all tested models, indicating that these models are highly effective in correctly identifying positive cases while avoiding incorrect classifications of negative cases. The TPR scale values show that the models are getting closer to 1. These results demonstrated that the LR model is still thriving despite being the poorest of all the models, even if its FPR is the highest at 14.05 percent.
- By comparing the results of the tested models with the results of some ML models included in Table 3, it was found that each of the proposed models was the most efficient among its peers.
- Logistic regression (LR) often performs less accurately on complex datasets such as NSL-KDD due to a drawback of the following:
  - ✓ Logistic regression, a linear classifier, struggles with complex, nonlinear data patterns. More flexible, nonlinear decision boundaries are often required to accurately classify attack types and network behaviors in the NSL-KDD dataset.
  - ✓ The NSL-KDD dataset's numerous characteristics interact in intricate ways, making logistic regression less adept at handling these interactions than nonlinear models like XGB or neural networks, which can learn intricate feature correlations.
  - ✓ Logistic regression's effectiveness may be hindered by its inability to adjust to irregular class distributions without additional steps like class weighting or resampling techniques.
  - ✓ Good feature engineering is crucial for logistic regression, but with advanced models like deep learning, intensive preprocessing is less needed, allowing for more intricate pattern recognition.

Conclusion

Smart cities are revolutionizing cyberattack prevention, utilizing machine learning-based intrusion detection strategies. IoT networks, which can be implemented in various industries, face challenges from hackers who can quickly disrupt network equilibrium.

This research develops machine-learning approaches to identify potential cyberattacks using the NSL-KDD dataset. The study integrates intrusion detection systems of anomaly, signature, and hybrid types in IoT networks. Four machine learning methods are used: XG Boost, Random Forest, KNN, and Logistic Regression.

According to the study, a new data set including nonlinear characteristics, neural networks, and ensemble approaches is required to identify intrusions on IoT networks and smart city challenges. Security countermeasures are also examined in the study. As the number of IoT devices increases, it is beneficial to employ supervised machine learning, deep learning, computational intelligence, optimization, genetic algorithms, supervised learning, reinforcement, and sliding mode controllers in intrusion detection systems. Network security researchers also require new datasets. Scientists should concentrate on real-time intrusion detection systems and secure network communication through new security-based routing techniques. IoT advancements, such as AI for predictive analytics and blockchain for secure data sharing, could shape the next generation of smart cities.

		Predicted Condition	
Total Pop. = P+N		Pos. (PP)	Neg. (PN)
Actual Cond.	Pos. (P)	TP	FN
	Neg. (N)	FP	TN
Total Pop. = P+N		(P)	(N)

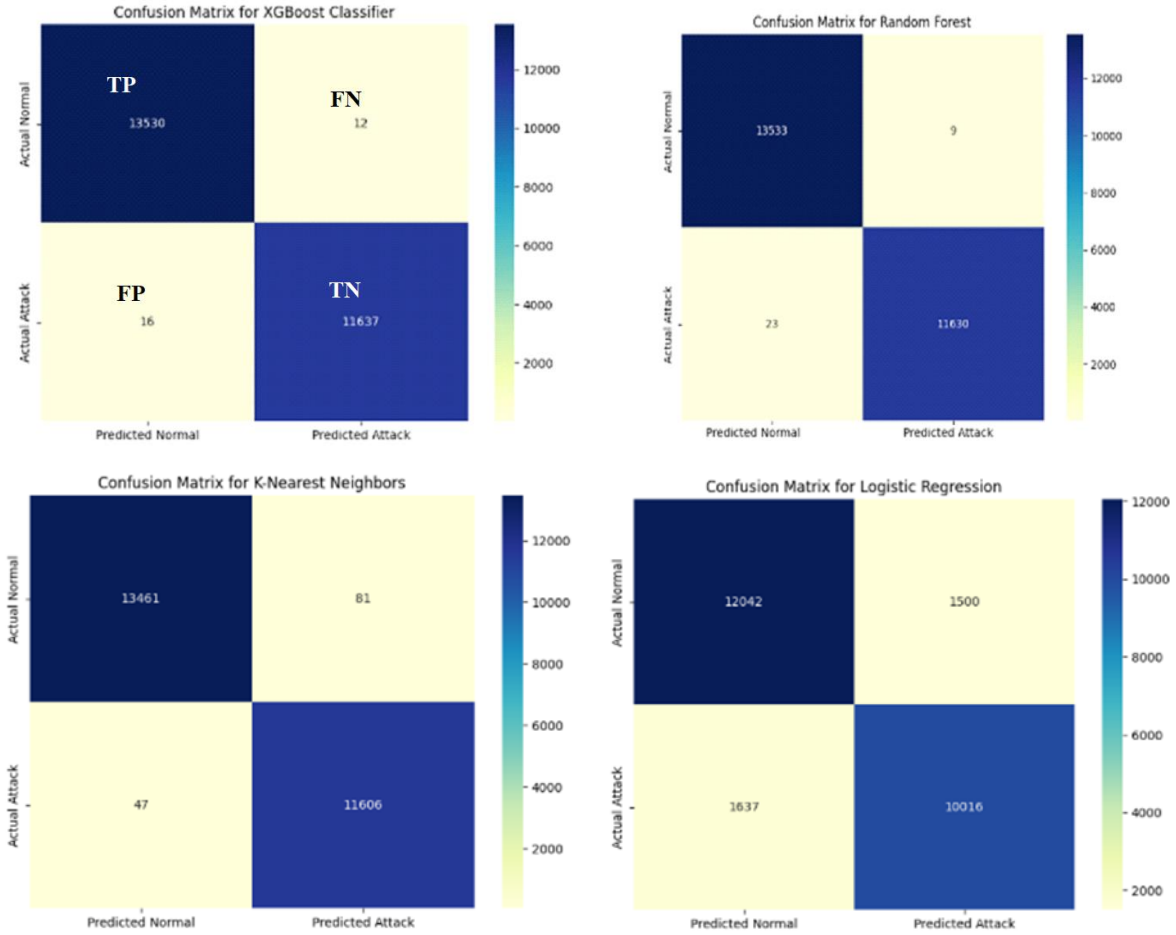

Figure 9. Confusion Matrices of ML models

Table 2. The Performance Values (%) of ML Models

ML Models	Acc.	Prec.	Recall	F1-sc.	TPR	FPR
XGBoost (XGB)	99.92	99.88	99.91	99.90	99.91	0.14
Random Forest (RF)	99.86	99.83	99.93	99.88	99.93	0.20
K-Nearest Neighbors (KNN)	99.57	99.65	99.40	99.53	99.40	0.40
Logistic Regression (LR)	88.05	88.03	88.92	88.48	88.92	14.05

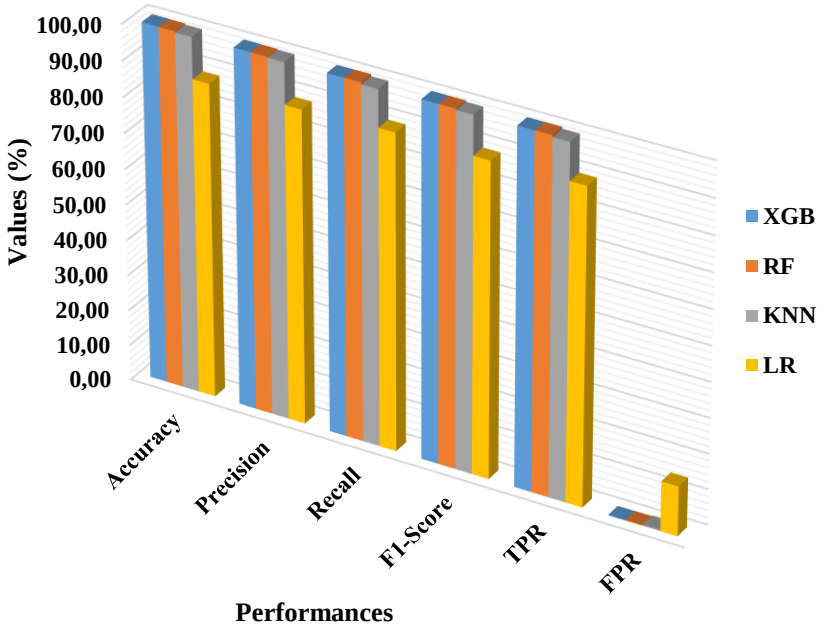

Figure 10. The Different Performance Values of ML Models

Table 3. A Comparison between different ML models based on NSL-KDD Dataset

Ref. No.	Model	Acc. (%)	Model Type
Proposed	XGBoost	99.92	Machine Learning (ML) Models
(Zhou & Paffenroth, 2017)	XGB	98.25	
(Hazem, 2023)	XGB	97.67	
(Priyadarshini, 2024)	XGB	97.06	
Proposed	Random Forest (RF)	99.86	
(Ou et al., 2012)	RF	98.6	
(Alhanaya & Al-Shqeerat, 2023)	RF	99.70	
(Priyadarshini, 2024)	RF	97.84	
Proposed	KNN	99.57	
(Zhou & Paffenroth, 2017)	KNN	98.31	
(Alhanaya & Al-Shqeerat, 2023)	KNN	99.10	
Proposed	Logistic Regression (LR)	88.05	
(Kim et al., 2020)	LR	78.0	
(Priyadarshini, 2024)	LR	68.97	

References

Abdullah, H. S. A. (2024). A comparison of several intrusion detection methods using the NSL-KDD dataset. Wasit Journal of Computer and Mathematics Science, 3(2).

Alars, E. S. A., & Kurnaz, S. (2024). Enhancing network intrusion detection systems with combined network and host traffic features using deep learning: deep learning and IoT perspective. Discover Computing, 27(1), 39.

Alhanaya, M., & Al-Shqeerat, K. (2023). Developing an integrated framework for securing internet of things traffic in smart cities using machine learning techniques. Applied Sciences, 13(16), 9476.

Al-Kasassbeh, M., Abbadi, M. A., & Al-Bustanji, A. M. (2020). LightGBM algorithm for malware detection. In Intelligent Computing: Proceedings of the 2020 Computing Conference, Volume 3 (pp. 391-403). Springer International Publishing.

Asadullah, M., & Raza, A. (2016, November). An overview of home automation systems. In 2016 2nd international conference on robotics and artificial intelligence (ICRAI) (pp. 27-31). IEEE.

Ashraf, J., Keshk, M., Moustafa, N., Abdel-Basset, M., Khurshid, H., Bakhshi, A. D., & Mostafa, R. R. (2021). IoTBoT-IDS: A novel statistical learning-enabled botnet detection framework for protecting networks of smart cities. Sustainable Cities and Society, 72, 103041.

Ayub, M. Y., Haider, U., Haider, A., Tashfeen, M. T. A., Shoukat, H., & Basit, A. (2023). An intelligent machine learning based intrusion detection system (IDS) for smart cities networks. EAI Endorsed Transactions on Smart Cities, 7(1), e4-e4.

Chen, L., Kuang, X., Xu, A., Suo, S., & Yang, Y. (2020, December). A novel network intrusion detection system based on CNN. In 2020 eighth international conference on advanced cloud and big data (CBD) (pp. 243-247). IEEE.

Chohan, M. N., Haider, U., Ayub, M. Y., Shoukat, H., Bhatia, T. K., & Ul Hassan, M. F. (2023). Detection of Cyber Attacks using Machine Learning based Intrusion Detection System for IoT Based Smart Cities. EAI Endorsed Transactions on Smart Cities, 7(2).

Cimen, H., Palacios-Garcia, E. J., Kolaek, M., Cetinkaya, N., Vasquez, J. C., & Guerrero, J. M. (2020). Smart-building applications: Deep learning-based, real-time load monitoring. IEEE Industrial Electronics Magazine, 15(2), 4-15.

Douiba, M., Benkirane, S., Guezzaz, A., & Azrour, M. (2023). An improved anomaly detection model for IoT security using decision tree and gradient boosting. The Journal of Supercomputing, 79(3), 3392-3411.

Fiore, S., Elia, D., Pires, C. E., Mestre, D. G., Cappiello, C., Vitali, M., ... & Aloisio, G. (2019). An integrated big and fast data analytics platform for smart urban transportation management. IEEE access, 7, 117652-117677.

Gaber, T., El-Ghamry, A., & Hassanien, A. E. (2022). Injection attack detection using machine learning for smart IoT applications. Physical Communication, 52, 101685.

Garcia-Teodoro, P., Diaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomalybased network intrusion detection: Techniques, systems and challenges. computers & security, 28(1-2), 18-28.

Gaur, A., Scotney, B., Parr, G., & McClean, S. (2015). Smart city architecture and its applications based on IoT. Procedia computer science, 52, 1089-1094.

Ge, M., Syed, N. F., Fu, X., Baig, Z., & Robles-Kelly, A. (2021). Towards a deep learningdriven intrusion detection approach for Internet of Things. Computer Networks, 186, 107784.

Gu, J., Wang, L., Wang, H., & Wang, S. (2019). A novel approach to intrusion detection using SVM ensemble with feature augmentation. Computers & Security, 86, 53-62.

Guezzaz, A., Asimi, A., Asimi, Y., Tbatou, Z., & Sadqi, Y. (2017). A lightweight neural classifier for intrusion detection. General Letters in Mathematics, 2(2), 57-66.

Guezzaz, A., Asimi, A., Mourade, A., Tbatou, Z., & Asimi, Y. (2020). A multilayer perceptron classifier for monitoring network traffic. In Big Data and Networks Technologies 3 (pp. 262-270). Springer International Publishing.

Guezzaz, A., Azrour, M., Benkirane, S., Mohy-Eddine, M., Attou, H., & Douiba, M. (2022). A Lightweight Hybrid Intrusion Detection Framework using Machine Learning for Edge-Based IIoT Security. Int. Arab J. Inf. Technol., 19(5), 822-830.

Guezzaz, A., Benkirane, S., Azrour, M., & Khurram, S. (2021). A reliable network intrusion detection approach using decision tree with enhanced data quality. Security and Communication Networks, 2021(1), 1230593.

Hazem S. Abdullah, "Artificial intelligence method for cyber security intrusion detection", MSc. Thesis, Comp. Sc. Dep. AUL, Nov 2023.

Hazman, C., Guezzaz, A., Benkirane, S., & Azrour, M. (2023). lIDS-SIoEL: intrusion detection framework for IoT-based smart environments security using ensemble learning. Cluster Computing, 26(6), 4069-4083.

Hodo, E., Bellekens, X., Hamilton, A., Dubouilh, P. L., Iorkyase, E., Tachtatzis, C., & Atkinson, R. (2016, May). Threat analysis of IoT networks using artificial neural network intrusion detection system. In 2016 International Symposium on Networks, Computers and Communications (ISNCC) (pp. 1-6). IEEE.

Jan, S. U., Ahmed, S., Shakhov, V., & Koo, I. (2019). Toward a lightweight intrusion detection system for the internet of things. IEEE access, 7, 42450-42471.

Kasongo, S. M. (2021). An advanced intrusion detection system for IIoT based on GA and tree based algorithms. IEEE Access, 9, 113199-113212.

Kim, W. S., Lee, W. S., & Kim, Y. J. (2020). A review of the applications of the internet of things (IoT) for agricultural automation. Journal of Biosystems Engineering, 45, 385400.

Liao, H. J., Lin, C. H. R., Lin, Y. C., & Tung, K. Y. (2013). Intrusion detection system: A comprehensive review. Journal of network and computer applications, 36(1), 16-24.

Mohammad, N., Muhammad, S., Bashar, A., & Khan, M. A. (2019). Formal analysis of human-assisted smart city emergency services. Ieee Access, 7, 60376-60388.

Mohy-Eddine, M., Guezzaz, A., Benkirane, S., & Azrour, M. (2023). An effective intrusion detection approach based on ensemble learning for IIoT edge computing. Journal of Computer Virology and Hacking Techniques, 19(4), 469-481.

Ou, Q., Zhen, Y., Li, X., Zhang, Y., & Zeng, L. (2012, June). Application of internet of things in smart grid power transmission. In 2012 third FTRA international conference on mobile, ubiquitous, and intelligent computing (pp. 96-100). IEEE.

Pai, V., & Adesh, N. D. (2021). Comparative analysis of machine learning algorithms for intrusion detection. In IOP Conference Series: Materials Science and Engineering (Vol. 1013, No. 1, p. 012038). IOP Publishing.

Panagiotis, F., Taxiarxchis, K., Georgios, K., Maglaras, L., & Ferrag, M. A. (2021). Intrusion detection in critical infrastructures: A literature review. Smart Cities, 4(3), 1146-1157.

Pradhan, B., Bhattacharyya, S., & Pal, K. (2021). IoT-based applications in healthcare devices. Journal of healthcare engineering, 2021(1), 6632599.

Priyadarshini, I. (2024). Anomaly detection of iot cyberattacks in smart cities using federated learning and split learning. Big Data and Cognitive Computing, 8(3), 21.

Rincy N, T., & Gupta, R. (2021). Design and development of an efficient network intrusion detection system using machine learning techniques. Wireless Communications and Mobile Computing, 2021(1), 9974270.

Roopak, M., Tian, G. Y., & Chambers, J. (2020, January). An intrusion detection system against DDoS attacks in IoT networks. In 2020 10th annual computing and communication workshop and conference (CCWC) (pp. 0562-0567). IEEE.

Shah, S. A., Seker, D. Z., Rathore, M. M., Hameed, S., Yahia, S. B., & Draheim, D. (2019). Towards disaster resilient smart cities: Can internet of things and big data analytics be the game changers?. IEEe Access, 7, 91885-91903.

Sikder, A. K., Petracca, G., Aksu, H., Jaeger, T., & Uluagac, A. S. (2021). A survey on sensorbased threats and attacks to smart devices and applications. IEEE Communications Surveys & Tutorials, 23(2), 1125-1159.

Simpson, S. V., & Nagarajan, G. (2021). An edge based trustworthy environment establishment for internet of things: an approach for smart cities. Wireless Networks, 1-17.

Ullah, I., & Mahmoud, Q. H. (2021). Design and development of a deep learning-based model for anomaly detection in IoT networks. IEEe Access, 9, 103906-103926.

Wang, D., Ding, B., & Feng, D. (2020, September). Meta reinforcement learning with generative adversarial reward from expert knowledge. In 2020 IEEE 3rd

International Conference on Information Systems and Computer Aided Education (ICISCAE) (pp. 1-7). IEEE.

Yuan, X., Chen, J., Zhang, N., Fang, X., & Liu, D. (2021). A federated bidirectional connection broad learning scheme for secure data sharing in Internet of Vehicles. China Communications, 18(7), 117-133.

Zhou, C., & Paffenroth, R. C. (2017, August). Anomaly detection with robust deep autoencoders. In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 665-674).