Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 2 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i02.52671

PAPER

Evaluation and Detection of Cyberattack in IoT-Based Smart City Networks Using Machine Learning on the UNSW-NB15 Dataset

Mubashir Ali[1], Shahbaz Pervez[1], Seyed Ebrahim Hosseini[1](✉), Muhammad Kashif Siddhu[2]

[1]Whitecliffe, Auckland, New Zealand

[2]University of Faisalabad, Faisalabad, Pakistan

seyedh@whitecliffe.ac.nz

ABSTRACT

With the proliferation of Internet of Things (IoT) devices across various applications, for example, smart homes, drones, and healthcare, the security vulnerabilities have also increased, necessitating robust network intrusion detection systems (NIDS). This study focuses on the classification of cyberattacks, including denial of service (DoS), worms, and backdoor attacks, from normal network traffic using the UNSW-NB15 dataset. machine learning (ML) and deep learning (DL) models, such as decision tree (DT) classifier, K-nearest-neighbor (KNN) classifier, linear regression, linear support vector machine, logistic regression (LR), multi-layer perceptron (MLP), and random forest (RF), were employed for both binary and multi-class classification. Data preprocessing involved handling null values, one-hot encoding categorical variables, and normalizing numerical features. Feature selection was performed using the Pearson correlation coefficient method, reducing the dataset attributes significantly. The models demonstrated high accuracy in detecting anomalies, with the RF classifier achieving the highest accuracy of 98.64% for binary classification and notable performance across multi-class classifications. This study underscores the effectiveness of ML techniques in enhancing IoT network security and offers comprehensive insights.

KEYWORDS

Internet of Things (IoT), smart city, network intrusion detection, cyberattacks, decision tree (DT) classifier, machine learning (ML), deep learning (DL), UNSW-NB15 dataset, denial of service (DoS), multi-layer perceptron (MLP), worms, backdoor, K-nearest-neighbor (KNN) classifier, linear regression, linear support vector machine, random forest (RF), data preprocessing, feature selection, binary classification, logistic regression (LR), multi-class classification

1 INTRODUCTION

In its 2024 Global Risk Report, the World Economic Forum named online attacks as one of the biggest challenges to the global financial structure [1]. There are now

Ali, M., Pervez, S., Hosseini, S.E., Siddhu, M.K. (2025). Evaluation and Detection of Cyberattack in IoT-Based Smart City Networks Using Machine Learning on the UNSW-NB15 Dataset. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(2), pp. 157–170. https://doi.org/10.3991/ijoe.v21i02.52671

Article submitted 2024-10-04. Revision uploaded 2024-12-01. Final acceptance 2024-12-05.

more security flaws due to the growth of Internet of Things (IoT) devices in smart cities. Despite improving automation and management, these networked devices' large and frequently unsecure networks make them vulnerable to cyber assaults. In order to protect these networks from potential threats, it is necessary to design strong network intrusion detection systems (NIDS).

Smart cities utilize IoT technology to connect and synchronize various devices and appliances, enabling remote monitoring, automation, and control. However, these interconnected systems are susceptible to intrusions due to the presence of sensitive personal and corporate data. Anomaly detection emerges as a promising approach for identifying illicit activities within smart cities [2]. Extant literature, on the other hand, is primarily concerned with breaches pertaining to the IoT and does not sufficiently address the identification of abnormalities that are unique to smart home environments. This problem is made worse by the dearth of thorough data that truly captures the complexity of smart homes, which include people with different skill levels and a wide range of constantly changing device kinds [3].

Cybercriminals find IoT devices appealing for a variety of reasons, including low technical proficiency among consumers, a high frequency of insecure IoT devices, insufficient security configurations, poor control applications, and the significant value attached to digital assets that are easily accessible. The IoT is anticipated to have nearly 50 billion linked gadgets by 2023, which increases the attraction of the network for cybercriminals due to its rapid growth [4].

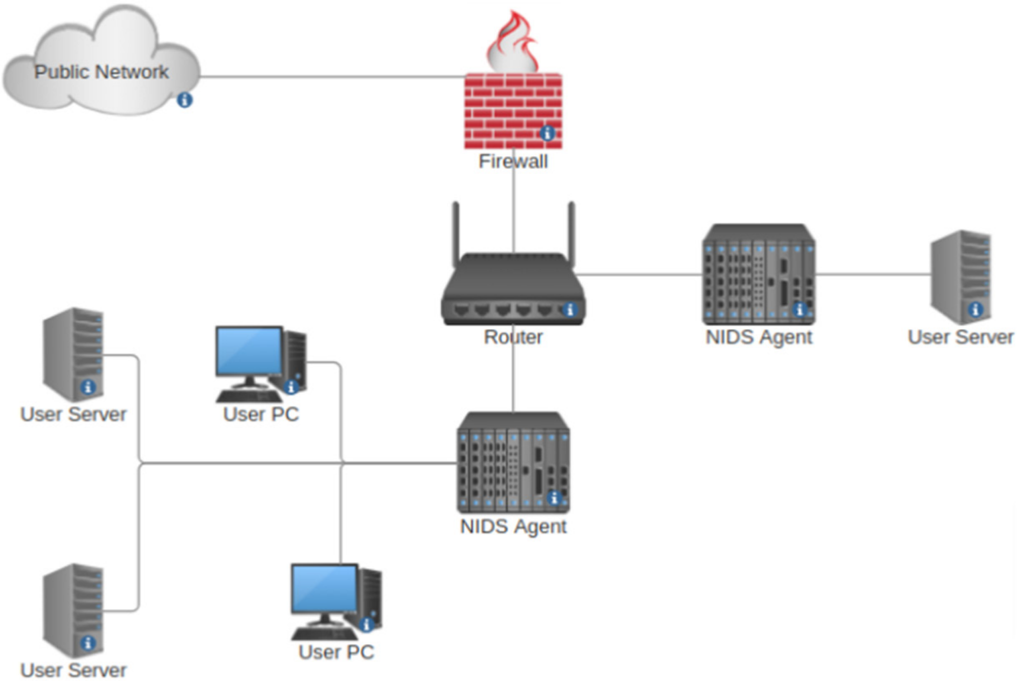

Fig. 1. Network intrusion detection system monitoring

The use of anomaly-based detection systems (ADS's) offers a chance to recognize unknown cyberthreats. Distinguishing between normal and deviant behaviors is the main goal inside the multidimensional ADS's realm. These systems need to first create a baseline of normal behavior to accomplish this. They must therefore continue to be on the lookout for any notable changes to the network or device that might point to questionable activity. It's especially challenging to stay relevant in this space because user behavior and cyberattack tactics are always changing [2].

The IoT connects more than just computers and smartphones to the internet. It encompasses a vast network of electronic devices, mechanical devices, and sensors. This expansion, while enabling exciting new applications, has also introduced significant security weaknesses. As the number of IoT devices continues to grow, so does the risk of cyberattacks [5].

The reach of IoT devices extends far beyond our traditional digital companions. From fire alarms and drones to smart home appliances and healthcare equipment, these interconnected devices are transforming numerous aspects of our lives. However, this growing network also presents a potential security nightmare. Malicious actors gaining access to these systems could have devastating consequences. To mitigate this risk, NIDS is deployed. These systems analyze network traffic, identifying and flagging malicious activity [6]. NIDS play a crucial role in protecting organizations' cloud, on-premises, and hybrid infrastructure, safeguarding them from cyberattacks targeting IoT devices.

The research addresses the challenge of detecting cyberattacks in IoT-based traffic within a smart city's network. It discusses the use of various machine learning (ML) models for classifying cyber intrusions, for example, worms, denial of service (DoS), and backdoor breaches, from typical network activity to intrusion detection. The study utilizes the UNSW-NB15 dataset for training these models. The key focus is on enhancing security measures for IoT devices deployed in smart cities due to the increased vulnerability posed by the expanding use cases of such devices. The absence of robust techniques to identify anomalies unique to smart home environments, compounded by a scarcity of comprehensive data reflecting the diverse makeup of smart homes, further underscores the significance of this study endeavor. Therefore, the research aims of the article can be encapsulated as follows:

- Develop a NIDS employing ML models to classify and detect various types of cyberattacks within IoT-based smart city network traffic utilizing the UNSW-NB15 dataset for training and evaluating.
- Prepare separate datasets for binary classification (normal vs. abnormal traffic) and multi-class classification (different categories of attacks) and evaluate and compare the performance of these models' employing metrics, for example, accuracy, mean absolute error, R2 score, mean squared error, and root mean squared error.

2 RELATED WORK

The UNSW-NB15 dataset has been used by numerous academics to assess the NIDS. For example, to approximate the radial basis function (RBF), in [7] kernel approximation technique used within a support vector machine (SVM). For their investigation, they integrated the NSL-KDD, Moore, and UNSW-NB15 datasets [8] leveraged the UNSW-NB15 and network information management and security group (NIMS) datasets to extract attributes relevant to domain name system (DNS), hypertext transfer protocol (HTTP), and MQTT attacks. They used artificial neural networks (ANN), decision trees (DT), and naive bayes (NB) as their three ML methodologies.

High-level features were extracted from the dataset using a deep feature embedding learning (DFEL) technique. SVM, gaussian naive bayes (GNB), K-nearest neighbors (KNN), gradient boosting tree (GBT), DT, and logistic regression (LR) are a few examples of ML techniques that were then applied for evaluation, as shown by [9]. The UNSW-NB15 dataset was used to create various numbers of clusters, and their effectiveness was assessed using Silhouette's measure [10], which verifies the consistency of data clusters. This was followed by assessment using several ML techniques based on DT, as shown by [11].

2.1 Cyberattacks

The distinct environments surrounding each smart home or similar setup make them susceptible to various logical cyberattacks and cyber threats emerging from computer networks and interconnected devices. Following the advancement of the Internet and computer networks, these cyber assaults have developed and adjusted, incorporating approaches such as DoS, sniffing, investigating, and so on [12]. The distinctive features of emerging IoT devices, including their limited computational capacity and a focus on prioritizing utility over security, make them particularly vulnerable to certain threats specifically designed for IoT. As integral components of the aforesaid environments, IoT devices can be exploited through any vulnerability. Many common cyberattacks, some dating back to the era of early communication technologies, continue to be applicable across various types of networks, involving IoT [12].

The cybercriminal exploits the attributes of the device to disrupt its ability to provide services, called DoS. By leveraging the computing capacity of the device and the bandwidth of the network, the attacker can render the device unusable, and its service availability is diminished. Smart home devices are especially susceptible to this type of attack due to their reduced processing capabilities and may even be affected by basic DoS attacks [13]. A distributed denial of service (DDoS) attack causes the server to become overloaded, causing it to become nonresponsive and reducing its service accessibility. The absolute most destructive instances of a DDoS attack was on the DNS Dyn, the service supplier, carried out by a botnet called Mirai, which consisted of infected smart home devices. This attack impacted numerous major websites, including Visa, PayPal, GitHub, and Amazon [14].

The primary goal of a worm (malicious software), once installed on a device, is to autonomously recreate and spread to connected equipment using the network. Consequently, the attacker overseeing this worm can impose significant damage on the intended recipient. Scholars have proposed that the worm could potentially spread by jumping from one compromised light source to nearby ones through their connections. Since Philips Hue lights haven't been fully verified, this intrusion could potentially escalate and propagate uncontrollably [15].

The backdoor technique is employed to bypass a device's security measures, such as authentication and encryption, to establish covert access and enable control without the user's knowledge. The study by [16] focuses on IoT backdoors, addressing issues such as weak cryptographic methods, evasion of system authentication, and exploitation of hardware vulnerabilities. This study highlights the associated risks and intrusion methods.

2.2 Intrusion detection systems

Intrusion detection systems (IDS) function as guardians in the digital world, constantly on the lookout for cyberattacks. They employ a diverse arsenal of strategies, network configurations, and techniques to identify these malicious attempts. To understand IDS capabilities, we can categorize them based on their focus (signature-based or anomaly-based), deployment location (network or host-based), and detection methods (behavioral or misuse-based). Evaluating their effectiveness relies on metrics that assess accuracy, precision, and overall usefulness in a real-world setting [17].

A NIDS detects cyberattacks using various methods. One type, known as misuse IDS or rule-based IDS, employs a database storing patterns and signatures of known attacks. It identifies a behavior as malicious if it matches any of these rules, necessitating constant updates. This method is similar to anomaly-based detection systems, but it relies on manually created standards that capture acceptable system behaviors. It detects deviations from these norms, which can be based on security policies. This approach enables an IDS to monitor internal node activity in IoT environments, ensuring that nodes comply with all routing protocols [18].

Network intrusion detection systems traditionally rely on signature-based detection, which struggles to identify novel attacks. ML offers a powerful alternative. By continuously analyzing network traffic, ML models can learn the patterns of normal activity. This allows them to detect deviations from these patterns, potentially uncovering new and unknown cyberattacks that traditional methods might miss. This integration of ML into NIDS enhances their ability to adapt and effectively safeguard networks against evolving cyber threats [19].

A typesetting team will do the final touch on your manuscript; you don't need to use specific styles to format it. However, the following points are the **minimum requirements**:

- File format: .doc or .docx
- Page size: A4
- Page margins: top/bottom: 52 cm; right/left: 4.4 cm
- Body text font size 10pt
- Equations should be editable (use the MS word formula editor)
- Correct citation style (see 2. citations/references)

Generally speaking, it should look like this document. Using the styles of this document may make things easier for you, but you don't need to use them.

2.3 Theoretical framework

This study delves into the critical domain of cyberattack detection within IoT-based smart city networks, firmly grounded in foundational concepts of network security. The study emphasizes the escalating security vulnerabilities accompanying the proliferation of IoT devices across various sectors. By employing a NIDS, organizations can comprehensively scrutinize network traffic to identify and thwart malicious activities. This approach underscores the paramount importance of safeguarding IoT ecosystems to guarantee the security and integrity of critical infrastructures. Leveraging diverse ML models, the research classifies a spectrum of cyberattacks ranging from DoS to backdoor infiltrations amidst normal network traffic. The use of the UNSW-NB15 Dataset for training and assessing models further fortifies the empirical foundation of the research. Through meticulous data preprocessing, feature engineering, and model selection, the study establishes a robust framework for enhancing intrusion detection capabilities within IoT environments. This empirical framework is underpinned by theoretical insights from cybersecurity, data science, and computer science disciplines, facilitating a multidimensional approach to cyber defense.

Our understanding of cyber defense mechanisms customized for IoT ecosystems in smart city contexts is advanced by the study's theoretical foundations, methodological rigor, and multidisciplinary approach. This all-encompassing method not only improves hack detection but also adds to the larger conversation about protecting vital infrastructures in the increasingly interconnected smart city environments.

3 METHOD

The research methodology employed in this study adopts a quantitative approach to investigate intrusion detection in the network traffic of IoT-based smart cities. The population under investigation comprises data on network traffic captured in the UNSW-NB15 Dataset, obtained from the cyber range lab of the Australian Centre for Cyber Security (ACCS). This dataset contains raw network packets and associated features, facilitating the analysis of various cyberattacks such as DoS, worms, backdoor, and others.

3.1 Classification of anomaly detection

Depending on whether labelled data is available, three distinct classifications of abnormality identification can be established [20]. The Table 1 below elaborates on the three categories.

Table 1. Classification of anomaly detection

Classifier	Description
Supervised Anomaly Detection	Using data labeled as "normal" and "anomaly," a model is developed to ascertain whether a new instance is normal or abnormal.
Semi-Supervised Anomaly Detection	During model generation, only instances from typical classes are considered. An anomaly denotes a novel sample that defies classification as normal.
Unsupervised Anomaly Detection	The classification model can be built without relying on training data with labels.

$$d = \begin{cases} < t, normal\ (under\ threshold) \\ > t, anomaly\ (above\ threshold) \end{cases} \tag{1}$$

Current approaches. Existing approaches to cyberattack identification in IoT-based networks of smart city traffic typically involve the use of ML models trained on labeled datasets [21]. These models are tasked with classifying various types of cyberattacks, such as DoS, worms, backdoors, and others, from normal network traffic. The UNSW-NB15 Dataset is commonly utilized for training these models, providing a diverse set of network traffic data with labeled attack categories. Strengths of existing approaches include their ability to leverage labeled datasets for supervised learning, enabling the development of accurate classification models.

Possible approaches. Possible approaches to cyberattack identification in IoT-based network traffic of smart cities encompass a range of innovative strategies and methodologies. One promising avenue involves the exploration of unsupervised learning techniques [22], which do not require labeled data for training and can thus adapt more easily to evolving threat landscapes. Another possible approach is the integration of anomaly detection algorithms, which can identify deviations from normal behavior in network traffic without relying on predefined attack signatures [23]. This approach may be particularly effective for detecting previously unseen attacks or zero-day exploits.

3.2 Data explanation and preprocessing

The dataset comprised 175,341 rows and 45 characteristics. The dataset contained 81,173 rows and 45 attributes after null values were removed. Datatype information from the features dataset is used to transform different data types for attributes.

Using pd.get_dummies(), the categorical columns "proto," "service," and "state" are one-hot encoded; these three characteristics are then deleted. 19 attributes made up the data_cat Dataframe following one-hot encoding. The primary dataframe and data_cat are concatenated. The dataframe's total attributes are 61. In relation to data normalization the MinMax scaler is used to scale the DataFrame's 58 numerical columns. The characteristics that exhibited a correlation coefficient of greater than 0.3 with the target attribute label were chosen.

The dataset was split into two categories for binary classification: 'normal' and 'abnormal' categories, while for multi-class classification, attacks were categorized into nine classes. Feature selection was performed using the Pearson Correlation Coefficient [24] method to identify attributes highly correlated with the target labels. Data analysis involved interpreting the performance of the models and comparing their effectiveness in detecting different types of cyberattacks from network traffic. The results obtained from the trained models were analyzed to draw meaningful conclusions regarding the efficacy of ML approaches in cyberattack identification for IoT-based smart city networks.

3.3 Binary classification

To classify any incoming network data as either normal or containing an attack, the models were trained on this binary categorization. Although this is a more straightforward method, it lacks information regarding the nature of the attack. The "label" property is divided into "normal" and "abnormal" categories. The 'label' property is saved with its encoded labels once it has been encoded using LabelEncoder(). Binary dataset with 61 columns and 81173 rows. Following feature selection, 'bin_data' has 15 properties, which are 'rate,' 'sttl,' 'sload,' 'dload,' 'ct_srv_src,' 'ct_state_ttl,' 'ct_dst_ltm,' 'ct_src_dport_ltm', 'ct_dst_sport_ltm', 'ct_dst_src_ltm', 'ct_src_ltm', 'ct_srv_dst', 'state_CON', 'state_INT', and 'label'.

3.4 Multi-class classification

The models are able to detect malicious traffic and identify the exact kind of attack that is being attempted. This gives security staff more precise information so they can respond appropriately. The 'attack_cat' characteristic is divided into nine groups. "Analysis," "Backdoor," "DoS," "Exploits," "Fuzzers," "Normal," "Generic," "Reconnaissance," and "Worms" Attack_cat is one-hot-encoded, and its labels are kept in label after being encoded using LabelEncoder().

Multiple classes Dataset: 69 columns, 81173 rows. Following feature selection, 'multi_data' has 16 properties, which are 'dttl', 'swin', 'dwin', 'tcprtt', 'synack', 'ackdat', 'label', 'proto_tcp', 'proto_udp', 'service_dns', 'state_CON', 'state_FIN', 'attack_cat_Analysis', 'attack_cat_DoS', 'attack_cat_Exploits', 'attack_cat_Normal'.

3.5 Approaches for validation

We utilized IDS validation techniques to ascertain the accuracy of a NIDS model in representing the system's ability to detect attacks. These methods include theoretical, empirical, and speculative procedures, which are employed to verify the effectiveness of NIDS [2]. Various criteria can be used to assess a NIDS, each of which should be applied based on its specific context. When constructing NIDS, essential criteria such as minimizing false alarms, managing resource usage efficiently, and effectively identifying threats in real-time need to be considered [2].

Various ML models were trained and evaluated on the preprocessed datasets, including DT classifier, KNN classifier, linear, RF classifier, multi-layer perceptron (MLP) classifier, linear regression model, and LR model. These are assessed using metrics such as mentioned underneath.

$$Accuracy = \frac{No.of\ Correct\ Predictions}{Total\ No.of\ Predictions} \times 100 \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

3.6 Analysis of the ML algorithms used

Support vector machines, or SVMs, are excellent at identifying the hyperplanes in data that best distinguish legitimate IoT traffic from malicious information, which makes them useful for identifying anomalies in IoT devices with limited resources.

Several DTs are used in RF, a potent ensemble technique, to provide reliable IoT attack detection. In order to detect patterns that differ from typical behavior, it analyzes a variety of network traffic features, which successfully flags possible assaults [25].

One such straightforward method for detecting IoT attacks is KNN. It uses the training data's normal traffic patterns and labelled attack patterns to determine how to classify new data pieces [26]. DTs are well-suited for IoT attack detection as they can learn a series of rules based on various data features (e.g., device type, data volume) to classify incoming traffic as normal or belonging to a specific attack category [27]. LR, unlike linear regression, can handle the binary nature of attack detection (normal vs. attack) in IoT networks. It analyzes various features of network traffic and predicts the probability of an event being an attack, offering a more nuanced approach than simple thresholds [25].

By passing IoT data through several layers, the MLP, a kind of ANN, can learn intricate patterns. This makes it useful for identifying non-linear correlations between features and possible IoT assaults [28]. Linear regression for detecting IoT attacks involves modeling the relationship between network traffic features and the likelihood of an attack to identify anomalies [29]. By comparing and assessing the performance of different ML algorithms, this study determined the best strategy for accurate and efficient hand signal detection in the context of PSL.

4 EVALUATION OF RESULTS AND DISCUSSION ON IMPACTS

Evaluation metrics such as accuracy, MAE, MSE, RMSE, and R2 Score assess the performance of ML algorithms. Accuracy gives an overall measure of correctness, but it might not be sufficient when dealing with imbalanced datasets. Precision focuses on the accuracy of positive predictions, while recall highlights the ability of the model to capture positive instances [30].

4.1 Results

In Table 2, among the ML models tested for IoT attack detection on UNSW-NB15, RF excelled with the highest accuracy (98.64%) and the best metrics (MAE, MSE, RMSE, R2). MLP followed closely, suggesting both are strong choices for intrusion detection in IoT.

Table 2. Binary classification evaluation

Classifier	Accuracy	MAE	MSE	RMSE	R2 Score
DT	98.09%	0.019	0.019	0.138	89.56
KNN	98.31%	0.017	0.017	0.130	90.74
Lin. Regression	97.81%	0.022	0.022	0.148	88.21
Lin. SVM	97.85%	0.021	0.021	0.147	88.45
Logistic Regression	97.80%	0.022	0.022	0.148	88.18
MLP	98.37%	0.016	0.016	0.128	91.11
RF	98.64%	0.014	0.014	0.116	92.60

Table 3 presents evaluation results for multi-class classification on the UNSW-NB15 dataset. The linear regression model achieves 95.13% accuracy and a notable R2 score of 91.82. Other models, including DT, KNN, Linear SVM, LR, MLP, and RF classifier, achieve accuracies between 97.20% and 97.59%. Linear SVM and LR stand out with the highest accuracy of 97.59%, MAE of 0.060, RMSE around 0.424–0.425, and R2 scores of 87.88-87.93. The RF classifier, with 97.32% accuracy, shows slightly higher MAE (0.066), RMSE (0.446), and R2 score of 86.64. Overall, linear SVM and LR demonstrate balanced performance in IoT attack detection, combining high accuracy, low error rates, and strong R2 scores.

Table 3. Multi-class classification evaluation

Classifier	Accuracy	MAE	MSE	RMSE	R2 Score
DT	97.20%	0.068	0.205	0.453	86.18
KNN	97.37%	0.065	0.194	0.441	86.93
Lin. Regression	95.13%	0.068	0.121	0.349	91.82
Lin. SVM	97.59%	0.060	0.179	0.424	87.93
Logistic Regr	97.59%	0.060	0.181	0.425	87.88
MLP	97.54%	0.061	0.179	0.423	87.98
RF	97.32%	0.066	0.199	0.446	86.64

4.2 Impact analysis

Cyberattacks on IoT equipment in smart cities can have a major negative effect on the social, technological, environmental, and economic spheres, among others. This section will examine these effects and highlight the significance of practical fixes like the one this paper offers.

- Social impact: Cyberattacks against vital infrastructure, such as power grids or transportation networks, can put public safety in jeopardy. Regular cyberattacks have the potential to reduce public confidence in smart city programs and make people reluctant to adopt them. Identity theft and privacy violations may result from cyberattacks that reveal personal data gathered by IoT devices.
- Economic impact: Cyberattacks have the potential to compromise vital services and infrastructure in smart cities, costing residents and companies' money. Costs related to data breaches, service interruptions, repairs, and lost productivity may fall under this category. Due to security risks and possible financial losses, the frequency of cyberattacks may deter private enterprises from investing in smart city infrastructure.
- Technological impact: Cyberattacks have the ability to overburden network resources and impede device connectivity, especially DDoS attacks. Cyberattacks may occasionally cause harm to or disable IoT devices. Strong security measures must be developed and implemented in response to the increased threat of cyberattacks; this will cost extra money and require continuous maintenance.

4.3 Limitations and challenges

While UNSW-NB15 is a valuable resource for network intrusion detection, its data may not fully reflect the realities of IoT-based smart city networks. The dataset focuses on traditional network traffic patterns, which might differ significantly from the communication protocols and data formats used by the diverse devices in a smart city. Additionally, the range of cyberattacks it captures might be limited to traditional network attacks, potentially missing newer methods that exploit vulnerabilities specific to IoT sensors or communication protocols. This mismatch between the data and the target environment can hinder the ability of ML models to accurately detect anomalies and cyberattacks in a smart city network.

Even if the data limitations are addressed, there are challenges inherent to deploying ML models for real-time protection of smart city networks. These models can be computationally expensive to train and run. The resource-constrained nature of many IoT devices, with limited processing power and battery life, may not be able to handle such models effectively. Furthermore, the dynamic nature of smart city networks, with ever-evolving configurations and potential for new attack methods, necessitates models that can continuously adapt and learn. This requires ongoing data collection, retraining, and potentially complex infrastructure to manage the process. These challenges need to be addressed to ensure the smooth operation and effectiveness of deep learning (DL) and ML-based security systems in smart city networks.

The study explores using ML models for cyberattack identification in IoT-based smart city network traffic. While the accuracy achieved is promising, there are several opportunities for future research to improve the effectiveness and applicability of these systems in real-world scenarios using deep learning.

- Continuously evolving threats: Cyberattacks are constantly developing new techniques to bypass IDSs. ML models can be enhanced to perform online learning, allowing them to adapt to evolving threats in real-time. This could involve incorporating anomaly detection techniques that can identify novel attack patterns without prior training data.
- Explainable AI for intrusion detection: ML models, especially DL models, can be opaque in their decision-making process. This lack of interpretability can hinder trust and adoption in safety-critical applications like intrusion detection. Research on explainable AI (XAI) methods can be applied to make the decision-making process of ML models in IDSs more transparent. This would allow security analysts to understand why the system flags certain traffic as malicious and improve overall system confidence.
- Scalability and resource efficiency: Deploying IDSs on resource-constrained IoT devices can be challenging due to limited processing power and battery life. Research on developing lightweight and efficient ML models specifically designed for IoT devices is crucial. This could involve exploring techniques such as model pruning, quantization, and knowledge distillation to reduce the computational complexity of models without sacrificing accuracy.

By addressing these future research directions, IoT-based network IDS can become more robust, adaptable, and trustworthy, ultimately leading to a more secure and resilient smart city infrastructure.

5 CONCLUSION

This study investigated the productivity of various ML models for cyberattack detection in IoT-based smart city network traffic. The UNSW-NB15 dataset was used to train and evaluate the performance of these models. Our findings demonstrate that all the explored models achieved high accuracy in both binary (normal vs. abnormal traffic) and multi-class (classification of different attack types) classification tasks. RF emerged as the most effective model.

These outcomes demonstrate the possibility of ML for developing robust and accurate NIDS for smart city networks. Nonetheless, it's critical to recognize some of this study's shortcomings. Firstly, the research relied on a single dataset (UNSW-NB15). Future work should explore the generalizability of these findings using additional datasets encompassing a wider variety of attack types and network configurations. Secondly, the computational demands of DL models can be significant. For deployment on resource-constrained IoT devices, further research is needed to develop lightweight and efficient models that maintain high accuracy. Overall, this study provides valuable insights into the potential of ML for securing IoT-based smart city networks. By addressing the limitations identified here, future research can pave the way for the development of practical and scalable NIDS solutions that can safeguard these critical infrastructures.

6 REFERENCES

[1] H. R. Watch, *World Report 2024: Events of 2023*. Seven Stories Press, 2024.

[2] A. Khraisat and A. Alazab, "A critical review of intrusion detection systems in the internet of things: Techniques, deployment strategy, validation strategy, attacks, public datasets and challenges," *Cybersecurity*, vol. 4, pp. 1–27, 2021. https://doi.org/10.1186/s42400-021-00077-7

[3] S. Singh, Q. Z. Sheng, E. Benkhelifa, and J. Lloret, “Guest editorial: Energy management, protocols, and security for the next-generation networks and Internet of Things,” *IEEE Trans. Ind. Informatics,* vol. 16, no. 5, pp. 3515–3520, 2020. https://doi.org/10.1109/TII.2020.2964591

[4] C. Bodei, P. Degano, G.-L. Ferrari, and L. Galletta, “Security metrics at work on the things in IoT systems,” in *From Lambda Calculus to Cybersecurity Through Program Analysis*, in Lecture Notes in Computer Science, A. Di Pierro, P. Malacaria, and R. Nagarajan, Eds., vol. 12065, 2020, pp. 233–255. https://doi.org/10.1007/978-3-030-41103-9_9

[5] N. Dhameliya, “Revolutionizing PLC systems with AI: A new era of industrial automation,” *American Digits: Journal of Computing and Digital Technologies,* vol. 1, no. 1, pp. 33–48, 2023.

[6] L. Shi, L. Wu, and Z. Guan, “Three-layer hybrid intrusion detection model for smart home malicious attacks,” *Computers & Electrical Engineering,* vol. 96, p. 107536, 2021. https://doi.org/10.1016/j.compeleceng.2021.107536

[7] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, “Shallow neural network with kernel approximation for prediction problems in highly demanding data networks,” *Expert Systems with Applications,* vol. 124, pp. 196–208, 2019. https://doi.org/10.1016/j.eswa.2019.01.063

[8] N. Moustafa, B. Turnbull, and K.-K. R. Choo, “An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things,” *IEEE Internet of Things Journal,* vol. 6, no. 3, pp. 4815–4830, 2018. https://doi.org/10.1109/JIOT.2018.2871719

[9] Y. Zhou, M. Han, L. Liu, J. S. He, and Y. Wang, “Deep learning approach for cyberattack detection,” in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2018, pp. 262–267. https://doi.org/10.1109/INFCOMW.2018.8407032

[10] N. Maryenko and O. Stepanenko, “Fractal dimension of silhouette magnetic resonance brain images as a measure of age-associated changes in cerebral hemispheres,” *Duzce Medical Journal,* vol. 25, no. 1, pp. 27–37, 2023. https://doi.org/10.18678/dtfd.1180625

[11] V. Kumar, A. K. Das, and D. Sinha, “Statistical analysis of the UNSW-NB15 dataset for intrusion detection,” in *Computational Intelligence in Pattern Recognition, Advances in Intelligent Systems and Computing,* A. Das, J. Nayak, B. Naik, S. Pati, and D. Pelusi, Eds., Springer, Singapore, vol. 999, 2020, pp. 279–294. https://doi.org/10.1007/978-981-13-9042-5_24

[12] F. Khan, R. Alturki, M. A. Rahman, S. Mastorakis, I. Razzak, and S. T. Shah, “Trustworthy and reliable deep-learning-based cyberattack detection in industrial IoT,” *IEEE Transactions on Industrial Informatics,* vol. 19, no. 1, pp. 1030–1038, 2022. https://doi.org/10.1109/TII.2022.3190352

[13] D. Geneiatakis, I. Kounelis, R. Neisse, I. Nai-Fovino, G. Steri, and G. Baldini, “Security and privacy issues for an IoT based smart home,” in *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2017, pp. 1292–1297. https://doi.org/10.23919/MIPRO.2017.7973622

[14] Z. Zhao *et al.*, “DDoS family: A novel perspective for massive types of DDoS attacks,” *Computers & Security,* vol. 138, p. 103663, 2024. https://doi.org/10.1016/j.cose.2023.103663

[15] J. I. I. Araya and H. Rifà-Pous, “Anomaly-based cyberattacks detection for smart homes: A systematic literature review,” *Internet of Things,* vol. 22, p. 100792, 2023. https://doi.org/10.1016/j.iot.2023.100792

[16] T. Abderrahmane, N. Amardjia, and T. Mohammed, “Securing laboratories through internet of things networks: A comprehensive approach for ensuring safety and efficiency,” *Bulletin of Electrical Engineering and Informatics,* vol. 13, no. 1, pp. 572–585, 2024. https://doi.org/10.11591/eei.v13i1.6728

[17] J. I. I. Araya and H. Rifà-Pous, "Anomaly-based cyberattacks detection for smart homes: A systematic literature review," *Internet of Things*, vol. 22, p. 100792, 2023. https://doi.org/10.1016/j.iot.2023.100792

[18] C. Hazman, A. Guezzaz, S. Benkirane, and M. Azrour, "Toward an intrusion detection model for IoT-based smart environments," *Multimedia Tools and Applications*, vol. 83, pp. 62159–62180, 2024. https://doi.org/10.1007/s11042-023-16436-0

[19] S. Abdulrezzak and F. Sabir, "An empirical investigation on Snort NIDS versus supervised machine learning classifiers," *Journal of Engineering*, vol. 29, no. 2, pp. 164–178, 2023. https://doi.org/10.31026/j.eng.2023.02.11

[20] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," *ACM computing surveys (CSUR)*, vol. 54, no. 2, pp. 1–38, 2021. https://doi.org/10.1145/3439950

[21] H. Makina and A. Ben Letaifa, "Bringing intelligence to Edge/Fog in Internet of Things-based healthcare applications: Machine learning/deep learning-based use cases," *International Journal of Communication Systems*, vol. 36, no. 9, p. e5484, 2023. https://doi.org/10.1002/dac.5484

[22] O. Iparraguirre-Villanueva, V. Guevara-Ponce, F. Sierra-Liñan, S. Beltozar-Clemente, and M. Cabanillas-Carbonel, "Sentiment analysis of tweets using unsupervised learning techniques and the k-means algorithm," *International Journal of Advance Computer Science and Applications (IJACSA)*, vol. 13, no. 6, 2022. https://doi.org/10.14569/IJACSA.2022.0130669

[23] M. Landauer, M. Wurzenberger, F. Skopik, W. Hotwagner, and G. Höld, "Aminer: A modular log data analysis pipeline for anomaly-based intrusion detection," *Digital Threats: Research and Practice*, vol. 4, no. 1, pp. 1–16, 2023. https://doi.org/10.1145/3567675

[24] C. Yuru, Z. Jing, L. Fan, L. Chuanxian, L. Fenglian, and L. Jinkui, "Correlation analysis of silicone oil deterioration index in cable termination based on pearson correlation coefficient method," in *2023 Panda Forum on Power and Energy (PandaFPE)*, 2023, pp. 1258–1262. https://doi.org/10.1109/PandaFPE57779.2023.10141368

[25] H. Samadi and M. A. Kollathodi, "A comprehensive comparison and analysis of machine learning algorithms including evaluation optimized for geographic location prediction based on Twitter tweets datasets," *Cogent Engineering*, vol. 10, no. 1, 2023. https://doi.org/10.1080/23311916.2023.2232602

[26] Q. Pan, M. Gao, P. Wu, J. Yan, and M. A. AbdelRahman, "Image classification of wheat rust based on ensemble learning," *Sensors*, vol. 22, no. 16, p. 6047, 2022. https://doi.org/10.3390/s22166047

[27] R. Kan, M. Wang, X. Liu, X. Liu, and H. Qiu, "An advanced artificial fish school algorithm to update decision tree for NLOS acoustic localization signal identification with the dual-receiving method," *Applied Sciences*, vol. 13, no. 6, p. 4012, 2023. https://doi.org/10.3390/app13064012

[28] J. C. Sekhar, J. Ramu, and V. K. Pratap, "Quantitative assessment of hand signal recognition using landmarks detection: A comparative study of machine learning techniques," in *2023 International Conference on Network, Multimedia and Information Technology (NMITCON)*, 2023, pp. 1–7. https://doi.org/10.1109/NMITCON58196.2023.10276054

[29] I. H. Sarker, "CyberLearning: Effectiveness analysis of machine learning security modeling to detect cyber-anomalies and multi-attacks," *Internet of Things*, vol. 14, p. 100393, 2021. https://doi.org/10.1016/j.iot.2021.100393

[30] R. Yacouby and D. Axman, "Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models," *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*, pp. 79–91, 2020. https://doi.org/10.18653/v1/2020.eval4nlp-1.9

7 AUTHORS

Mubashir Ali is with the Whitecliffe, Auckland, New Zealand (E-mail: mubashir4ali@yahoo.com).

Shahbaz Pervez is with the Whitecliffe, Auckland, New Zealand (E-mail: shahbazp@whitecliffe.ac.nz).

Seyed Ebrahim Hosseini is with the Whitecliffe, Auckland, New Zealand (E-mail: seyedh@whitecliffe.ac.nz).

Muhammad Kashif Siddhu is with the University of Faisalabad, Faisalabad, Pakistan (E-mail: kashifsiddhu215@gmail.com).