Alexandria Engineering Journal (2022) 61, 9395–9409

HOSTED BY

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

A machine learning-based intrusion detection for detecting internet of things network attacks

Yakub Kayode Saheed [a], Aremu Idris Abiodun [b], Sanjay Misra [c,*], Monica Kristiansen Holone [c], Ricardo Colomo-Palacios [c]

[a] School of IT & Computing, American University of Nigeria, Nigeria
[b] Department of Computer Science, Lagos State Polytechnic, Ikorodu, Nigeria
[c] Department of Computer Science and Communication, Østfold University College, Halden, Norway

Received 21 October 2021; revised 5 February 2022; accepted 27 February 2022
Available online 28 March 2022

KEYWORDS

Intrusion Detection System;
Machine Learning;
Internet of Things;
Min-max Normalization;
UNSWNB-15;
Principal Component Analysis;
Cat boost;
XgBoost

Abstract The Internet of Things (IoT) refers to the collection of all those devices that could connect to the Internet to collect and share data. The introduction of varied devices continues to grow tremendously, posing new privacy and security risks—the proliferation of Internet connections and the advent of new technologies such as the IoT. Various and sophisticated intrusions are driving the IoT paradigm into computer networks. Companies are increasing their investment in research to improve the detection of these attacks. By comparing the highest rates of accuracy, institutions are picking intelligent procedures for testing and verification. The adoption of IoT in the different sectors, including health, has also continued to increase in recent times. Where the IoT applications became well known for technology researchers and developers. Unfortunately, the striking challenge of IoT is the privacy and security issues resulting from the energy limitations and scalability of IoT devices. Therefore, how to improve the security and privacy challenges of IoT remains an important problem in the computer security field. This paper proposes a machine learning-based intrusion detection system (ML-IDS) for detecting IoT network attacks. The primary objective of this research focuses on applying ML-supervised algorithm-based IDS for IoT. In the first stage of this research methodology, feature scaling was done using the Minimum-maximum (min–max) concept of normalization on the UNSW-NB15 dataset to limit information leakage on the test data. This dataset is a mixture of contemporary attacks and normal activities of network traffic grouped into nine different attack types. In the next stage, dimensionality reduction was performed with Principal Component Analysis (PCA). Lastly, six proposed machine learning models were used for the analysis. The experimental results of our findings were evaluated in terms of validation data-

* Corresponding author.
E-mail addresses: yakubu.saheed@aun.edu.ng (Y. Kayode Saheed), sanjay.misra@hiof.no (S. Misra), monica.kristiansen@hiof.no (M. Kristiansen Holone), ricardo.colomo-palacios@hiof.no (R. Colomo-Palacios).

Peer review under responsibility of Faculty of Engineering, Alexandria University.

https://doi.org/10.1016/j.aej.2022.02.063

set, accuracy, the area under the curve, recall, F1, precision, kappa, and Mathew correlation coefficient (MCC). The findings were also benchmarked with the existing works, and our results were competitive with an accuracy of 99.9% and MCC of 99.97%.

1. Introduction

The Internet of Things (IoT) is a joint network of interconnected devices; these devices can decide without any human interventions. The advancement of various technology fields, like automatic identification, sensors, tracking, wireless communications, embedded computing, distributed services, and 5G networks, has increased the possibility of utilizing advanced objects in our daily activities via the Internet [1]. The IoT is defined by the intersection of the Internet and intelligent objects capable of communication and interaction. This new paradigm has been identified as a key player in the ICT business in the coming years [2]. In the IoT, a thing can be anything on the planet: a person with a blood pressure monitor implant, a car equipped with sensors that alert the driver when the tire pressure is low, a farm animal with a transponder, or any object that can be given an IP address and the ability to transfer data over a network [3]. According to Cisco, it is stated that by 2020, about 50 billion devices will be connected to the Internet [4]. Cisco Systems forecasted that the Internet of Things would generate $ 14.4 trillion in income and cost savings for businesses between 2013 and 2022 [5,6,7,8]. These connected devices – dubbed the IoT holds a lot of promise for improving social and corporate life as well as market development, increased accessibility necessitates the use of stronger security precautions [9]. The main reason for the network's poor performance is that it consumes a lot of energy due to its low battery capacity. As a result, reducing energy consumption is a critical requirement for achieving quality of service (QoS) in the IoT context. IoT devices could be healthcare devices, wearables, industrial robots, smart televisions, smart city infrastructures that can be monitored remotely. There are many interesting applications of IoT. Even if the IoT seems to be a more industrial phrase, about 87 percent of individuals still do not comprehend what it means [10].

There are two basic explanations for the majority of security problems and broad privacy concerns. The first is that IoT objects are constrained in terms of memory capacity, energy consumption [11], and processing capability [12]. Because of these limitations, traditional Internet approaches like the Advanced Encryption Standard (AES) and RSA [13] may be difficult to implement directly in the IoT [14]. End-to-end secure communications in rich-resource items like tablets, phones, and laptops, for example, can be achieved at the transport layer via Transport Layer Security (TLS) or at the network layer via Internet Protocol Security (IPsec). However, these approaches cannot be directly applied in limited-resource objects, and their absence could lead to eavesdropping, network side-channel attacks, and tracking, among other security and privacy threats. As a result, incorporating machine learning-based intrusion detection into the IoT paradigm is critical to combat these assaults while still meeting IoT criteria. The IoT is becoming more pervasive in our daily lives, connecting physical things to e-services. That is to say; the IoT is the engine that powers home automation, advanced manufacturing, modern health, and smart cities. Current technological advancements are propelling the creation of a connected knowledge-based society; our economies, communities, government machinery [15], and Critical National Infrastructure (CNI). Smart homes, intelligent transportation, health care, smart cities, and smart grids are all significantly reliant on IoT devices and smart technologies. While CNI concepts facilitate daily tasks, their reliance on IoT and ICT appliances introduces significant safety threats. Security with a high degree of severity attacks against CNI ideas, such as spoofing, data escape, denial of service, energy bleed, unsecure gateways, etc., are directed at sensitive information, can impair the availability of systems and energy supplies, resulting in system blackouts and other widespread and long-lasting harm. These security concerns may have a significant impact on the operation of services. For example, mass transit networks can be aimed to cause havoc all through peak travel time frames, attacks on power grids can result in massive energy waste and a possible system blackout, and thus require urgent attention [11]. Attack detection in IoT is fundamentally different than in the past due to the unique requirements for services of IoT that a central cloud cannot meet: resource constraint, distribution, low latency, mobility, and scalability, to name a few [16]. This implies that neither cloud-based nor standalone threat discovery technologies are enough to address the IoT's security issues. As a result, an intrusion detection system should be examined to bridge the gap [17]. It is becoming increasingly necessary to do continuous research in the intrusion detection domain in IoT networks. In terms of network security and more specifically in spotting intrusions, a key worry emerges as a result of the adoption of the IP protocol in version 6 (IPv6) because there is a link to the IoT IPv6 protocol [18]. This convergence of IPv6 and the IoT paradigm enables unrestricted Internet connectivity for a wide variety of products, including microwaves, blenders, clothes [19], cognitive buildings [20], and wearable gadgets [21]. Among other things, this makes network security a current issue, necessitating the development of IoT-specific intrusion detection algorithms.

Numerous studies have been conducted to control the optimal settings and consequences for intrusion detection in IoT environments [22,10]. The authors [23] demonstrate that detection is a critical activity since it identifies aberrant data within an assumed data collection. Diro et al., [24] conducted experiments on intrusion detection in computer networks using three (3) original datasets, namely KDDCUP '99, NSL-KDD, and ISCX. They suggested a distributed deep-learning system for detecting IoT/fog network attacks, and their studies have shown that artificial intelligence may be successfully applied to cybersecurity goals. Additionally, the authors devised and implemented an attack detection system in a dis-

persed system for IoT applications in smart cities. Using the ECDSA method, the authors [25] offer a key exchange protocol for cluster-based wireless sensor networks in an IoT environment that improves the probability of key exchange between sensor nodes and CH. It increases key management performance in terms of key sharing, but the computational complexity is substantially raised by the rekeying procedure. The study [26] suggested a hash key-based key management mechanism for cluster-based wireless sensor networks that employs random key pre-distribution for WSN in IoT. Three performance indicators were used in the study: packet loss rate, energy usage, and latency. Although the effort enhances performance by establishing a secure link for one-hop and multi-hop communication, the network's cluster heads are not movable. In 6LoWPAN and IEEE 802.15.4, the researchers [27] provided a system for detecting denial of service threats. For simulation, it is tested and built using the Ebbits network adapter platform and Contiki OS. The security manager component of Ebbits includes a DoS protection manager. It collects intrusion detection warnings via a network-based intrusion detection system built on Suricata, an open source intrusion detection system.

Previous research looked at solutions from the standpoints of IoT security threats and practices, as well as different machine learning algorithms, datasets, and implementation tools. Additionally, some have always focused on encryption and cryptography methods [28], which are dependent on key management techniques. Because keys are shared between sensor nodes, there are still key management difficulties in encrypted solutions [29]. Many key management systems [30] are probabilistic when it comes to key sharing in a clustered environment. A probabilistic key management method cannot guarantee that two nodes in separate clusters will be able to establish a shared key; if some of the neighbor nodes are unable to do so, they will be unable to participate in the network. Furthermore, because the same key is used by multiple nodes, the network is at higher risk if any of them is compromised. Hence, we proposed an ML-based IDS for detecting IoT network attacks. As a result, communication overhead is reduced, and there is no need for a foreign key among Cluster-Head (CH) and Cluster-Node (CN) for secret communication whenever the cluster node transfers to a new region. We investigate an ML-based IDS because ML-based models work well in enhancing scalability and minimizing energy consumption.

IoT devices are anticipated to develop more predominant than smartphones and have access to the most up-to-date complex information such as confidential information [31]. As a result, the number of attacks will increase, and the attack predictor variables will increase [32]. Another major challenge of IoT in the health care industry will be to provide network safety concerning possible attacks in health care systems [33]. IDS is a technology established to address network security. As a result of IoT applications' important application in our lives, it is important to advance IoT machine learning-based IDS capable of attacks detection.

The IoT is a novel cohort of IT, which is presently a hot topic of research for organizations, citizens, and governments around the world, with is security issues gaining more consideration [34]. IDS technology is a significant technique to safeguard the network's security, which is presently a popular topic in IoT security [35]. Hence, this paper aims to present IDS-based machine learning algorithms for detecting IoT attacks. The central contributions of this research are as follows.

- To use the Minimum-Maximum (Min-Max) normalization technique to ensure that all the feature values are on the same scale.
- To adopt PCA for dimensionality reduction to transform the data into principal components.
- To design and implement IDS that satisfy IoT protocol requirements with the UNSWNB-15 dataset as against dataset that suffers from numerous issues obtained in a conventional network.
- To develop several lightweight IDS models for IoT networks that are efficient.
- To compare the performance of the proposed models with existing techniques.

This paper is organized as follows. Section 2 discusses the related work. Section 3 presents the proposed methodology. Section 4 reports the results and discussion. Section 5 concludes the paper.

2. Related work

In this section, the past studies of IDS in IoT are presented. Diro et al. [24] proposed a distributed deep learning-based IoT attack detection system. The results of their work gave 96% accuracy. Hoda et al. [36] propose an IDS with low-capacity devices for IoT applications. The experimental results of their work achieved 99.4% for Denial of Service. The information of the dataset used for the analysis was not provided. The study of [37] utilized the NSL-KDD dataset with the deep learning (DL) method in cybersecurity. This work used a self-taught DL approach in which sparse-auto encoders were used to perform unsupervised feature learning on training data. The learned characteristics were used to classify the labeled test dataset into attack and normal. The authors evaluated performance using the n-fold cross-validation methodology, and the obtained result seemed to be reasonable. Two recent publications can also be referred to that focus on applying ML approaches to issues with security in IoT architectures using the KDD99 dataset.

Bostani and Sheikhan [38] introduced an altered K-means strategy for shrinking the training dataset and balancing the data used to train ELMs and SVMs. The experimental findings of the proposed model gave 96.02% percent accuracy and a 5.92 percent false alarm rate. The clustering with self-Organized Ant Colony Networks (CSOACN) was used to categorize network traffic as benign or usual traffic.

The authors [24] described attack detection using a fog-to-things design. The authors conducted a comparison between a shallow and a deep neural network utilizing a free online dataset.

The fundamental objective of this effort was to spot four distinct types of attack and abnormality. The system achieved 98.27 percent accuracy with a DNN model and 96.75 percent accuracy with a shallow NN model. According to the authors [38], the wireless sensor networks and Internet, respectively, IoT's primary units, are insecure, making the IoT suffer from various assaults. The same researchers suggest a new architec-

ture for real-time-based intrusion detection, consisting of the anomaly-based intrusion detection components and requirements for spotting two types of routing assaults referred to as selective and collectors routing assaults in the IoT. When the selective attack and collector attack were conducted concurrently, the suggested hybrid real-time technique produced a true-positive of 76.19% and a false positive of 5.92%.

Anthi et al., [11] proposed an IoT-based intrusion detection system. The authors successfully used several machine learning models to recognize network monitoring probing and simple kinds of DoS attacks for this purpose. The data set is generated by capturing network traffic for four (4) successive days with the software known as Wireshark. Weka was utilized to apply machine learning classifiers.

The study [39] proposed an intrusion detection model that uses a two-dimension reduction and classification module with two tiers. Additionally, this model was created to detect malicious behavior such as R2L and U2R attacks. The PCA and LDA were employed to decrease the dimensions. The entire experiment was conducted using the NSL-KDD dataset. NB and the Certainty Factor version of K-NN were used in the two-tier classification module to detect suspicious activities.

Kozik et al., [40] demonstrated a cloud-based classification-based threat detection solution. In this paper, an ELM scaled in the Apache Spark cloud infrastructure is used to analyze simulated Netflow structured data. The authors in [41] proposed different emerging technologies for the treatment, study, and investigation of victims with Covid-19. The result of their findings showed that technologies like Big Data, AI, and IoT would be essential for the treatment of sick-person with Covid-19. The work in [42] proposes a platform based on the IoT to identify and monitor Covid-19 occurrence. They used ML algorithms SVM, Decision Stump, K-NN, NN, Decision Table, NB, ZeroR, and OneR. The experimental results of their findings revealed that five (5) of these classification algorithms gave an accuracy of more than 90%. The study in [43] proposed a platform for healthcare professionals in the Covid-19 epidemic by utilizing artificial intelligence and the Internet of things. The authors found out that the challenges faced by health care workers can be reduced with the adoption of IoT. Table 1 depicts existing methods for IoT attacks classification using different ML strategies.

The majority of research in the literature focused on potential solutions from the perspectives of IoT standards and technologies, architecture types, IoT security threats, and practices, dissimilar machine learning approaches, datasets, and tools for implementation.

Unlike the previous efforts, in this paper, we study IDS approaches for resource-constrained devices in the network. The distinction is that the technique will perform feature scaling with the min–max method and do classification independently using its chosen features, which are chosen by PCA. Table 1 explains and summarizes the existing methods. The majority of the classifiers use ELM, K-means, LDA, and SVM algorithms for feature selection, as shown in the table. The majority of existing studies utilize the NSLKDD and KDDCUP99 datasets to conduct experiments. It is an older dataset in terms of identifying R2L, DoS, probing, and U2R assaults, whereas the UNSW-NB15 dataset, which is the most recent and captures a wireless network in terms of detecting exploits, DoS, generic, fuzzers, reconnaissance, backdoors, and worms, will be used for this research.

2.1. Motivation of the present work

The IoT is the driving force behind home automation, improved manufacturing, modern healthcare, and smart cities. Our businesses, communities, government machinery, and Critical National Infrastructure are all being pushed toward the formation of a linked knowledge-based society by CNI. IoT devices and smart technologies are critical in smart homes, intelligent transportation, health care, smart cities, and smart grids. In the sphere of IoT, anomaly and attack detection in the IoT ecosystem is a growing concern. Threats and attacks against IoT infrastructure are increasing in lockstep with the increasing utilization of IoT infrastructure across all domains. Thousands of assaults are known to emerge regularly as a result of the addition of multiple protocols, primarily from IoT. The majority of these assaults are minor variations of previously identified cyberattacks. This shows that even advanced techniques like cryptography have a hard time identifying even tiny mutations of threats with time. The success of ML in numerous big data sectors has sparked interest in cybersecurity. Because of improvements in CPU characteristics, the application of ML has been practicable. In this research, we have adopted the Min-max approach for feature scaling and PCA for FS. However, there are other feature scaling methods such as the z-score technique, and FS like LDA. The z-score technique necessitates the knowledge of the standard deviation, which isn't always possible while the LDA is sensitive to outliers. As a result, we decided to employ Min-max in the first phase and PCA in the second.

3. Proposed model

This section discusses the proposed IDS for detecting IoT applications attacks. The data generated from a smartphone and sensors are uploaded to the cloud. The data in the cloud is not safe as this data is vulnerable to attack. The attack can be launched directly on the cloud or via transmission. The dataset employed in this research is the UNSW-NB15 dataset [46]. This dataset contains up-to-date attack types and was released recently [47]. In the first stage of this research methodology, feature scaling was done using the min–max concept of normalization on the UNSWNB15 dataset to limit information leakage on the test data. In the next stage, dimensionality reduction was performed with PCA. Data preprocessing was the first analysis performed after the acquisition and loading of the dataset. Data preprocessing is very vital as it helps in eliminating outliers and removing redundant attributes. The Normalization method with the Min-max technique was used for data preprocessing. The output of the min–max is fed into the feature selection algorithm known as PCA. The PCA selected ten (10) important components out of the forty-nine attributes in the dataset. The reduced dataset is then trained by the XGBoost, CatBoost, KNN, SVM, QDA, and NB classifiers. The architecture of our proposed model is shown in Fig. 1.

3.1. Data preprocessing

Data preprocessing can be seen as a significant task in the ML field as it helps to eliminate defects in the dataset [48]. This is the first stage of the proposed methodology. It is projected to

Table 1 Existing methods for IoT attacks classification.

Research Paper	Year	Methodology	Results	Limitations
Niyaz et al., [37]	2015	Self-taught DL Sparse auto-encoder	STL: F-measure = 98.84%; SMR: F-measure = 96.79%	The dataset used was obtained in a traditional network setting and not suitable for IoT protocols.
Hodo et al., [36]	2016	ANN	99.4% accuracy	No information on the dataset used.
Bostani and Sheikhan [38]	2016	K-means with ELMs and SVMs	96.02% accuracy, 76.19% of TP rate, and 5.92 false rate	Low TP rate and high false alarm rate
Pajouh [44]	2016	Self-organized ant colony networks	Accuracy = 99.79 for DoS attack and accuracy = 98.55 for Probe attack	The dataset used does not reflect present-day attacks
Diro and N. Chilamkurti., [24]	2017	DNN and shallow NN models	Shallow NN = 96.75% accuracy; DNN = 98.27% accuracy	NSLKDD dataset was used which does not reflect present-day attacks
Anthi et al., [11]	2018	NB	Recall = 97.7% Precision = 97.7% and F-measure = 97.7%	The features of the dataset generated does not represent network behavior in a diverse environment.
Kozik et al., [40]	2018	ELM	83% accuracy	High training time
Pajouh et al., [39]	2019	LDA for dimensionality reduction with NB and CF-KNN for classification of network traffic	Accuracy = 84.82% and false alarm rate = 5.56	Low detection accuracy and high FP rate
Chen et al., [45]	2020	Decision tree	Accuracy = 99.98, precision = 97.38, recall = 97.39, F1 = 99.98	High training time to train the models

change the raw network IoT attack data to a format that is effective to use for further analysis [49].

3.1.1. Normalization

Normalization is a feature scaling technique in which the aim is to have all the values of the attributes on the same scale. There are various normalization methods, including standardized moment, z-score normalization, and min–max normalization [50]. We used the min–max normalization technique in this paper.

3.1.2. Min-max normalization technique

According to [51] attributes are normalized in the range [0,1] based on the equation.

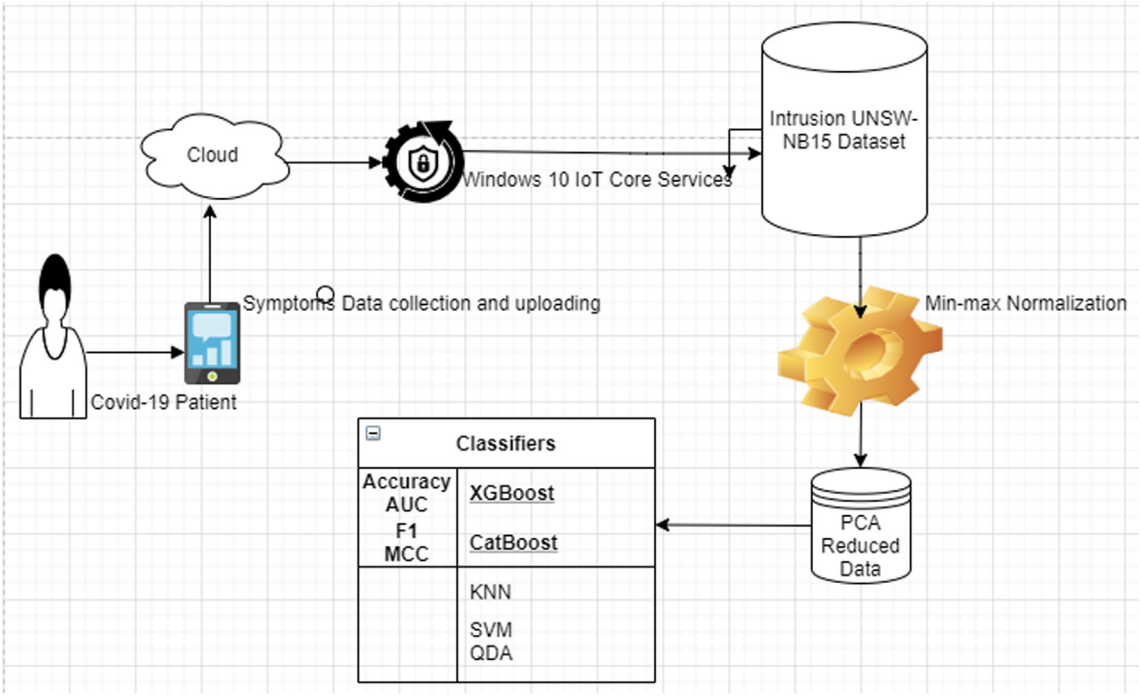

Fig. 1 Proposed model Architecture.

$$Z_{Norm} = Z - \min(x)/\max(z) - \min(z) \quad (1)$$

Here, min (z) and max (z) are the minimum, and the maximum values of the attribute Z. The original and the normalized value of the feature, Z, are indicated by Z and $Z_{Norm,}$ respectively.

3.2. Feature selection

The process of choosing features is referred to as feature selection [52] that contributes most to the prediction variable and reduces the training time and prevents computational complexity [53]. The feature selection technique is a dimensionality reduction approach that can be used to remove irrelevant attributes [54] in high dimensional data such as IDS data. The features contain information about the target, more feature means more information and better discriminative power [55]. However, this may not always hold that having more features does not mean more classification power. We used PCA for feature selection in this study.

3.2.1. Principal component analysis

The PCA is an extensively used unsupervised method for feature selection [56] and the oldest technique in multivariate statistical analysis [57]. The reason for utilizing PCA is to reduce the dimensionality by keeping the significant attributes information in the dataset [58]. It reduces the variable count using the orthogonal linear combinations with the significant variance [59]. PCA also selects the best important subset of the features in the dataset for classification [60]. In this paper, we use PCA to compress the attribute space where ten (10) components are selected to lower the dimensionality of the UNSW-NB15 dataset. The y_1 principal component is the linear mixture [61] of the main attributes with the maximum difference. The values of the first constituent are shown as equation (2):.

$$Y_1 = Lb_1 \quad (2)$$

M is the samples number, L is the matrix of observation with a mean of zero, and b_1 is the vector with the greatest variance (y), as expressed in equation (2);.

$$1/my_1 \ {}^1y_1 = 1/mb_1'U'Ua_1 = b_1Kb_1 \quad (3)$$

K is the covariances and variances matrix after the observations. To find solution of the equation (3), constraint $b_1\ b_1 = 1$ must be recognized utilizing Lagrange multiplier idea.

$$Z = b_1Kb - v(b_1b_1 - 1) \quad (4)$$

Maximizing Equation 4 entails deriving it in terms of its constituents b_1 until zero:.

$$\partial Z/\partial b_1 = 2Db_1 - 2vb_1 = 0 \quad (5)$$

The equation (5) gives in $Db_1 = vb_{1,}$ where b is an eigenvector of D, and v is the eigenvalue.

3.3. Extreme gradient boosting

Xgboost is one of the recently introduced ensemble machine-learning algorithms [62], highly effective tree boosting that has been used to generate state-of-the-art results in different applications. Xgboost uses the idea of ensembles of the tree to execute feature selection to select the feature importance [63]. The feature importance selected by the Xgboost in this paper is shown in Fig. 2.

3.4. Cat Boost

The Cat boost algorithm is a powerful machine learning technique that has been used to generate outstanding results in various applications [64]. Though, Cat Boost is developed to handle categorical features. However, it can still handle continuous or numerical attributes [65]. Cat boost model is a special feature that is added to the gradient-boosting decision tree algorithm [66]. The cat boost pseudocode can be shown in Table 2.

3.5. K Nearest neighbor

The KNN is one of the simplest classifiers in machine learning. The k-NN approach constructs a model of the target function from all labeled training cases. K-NN method to classification is completely non-parametric [67,68] and a method for classifying objects based on the nearest training samples in the feature space using instance-based learning. The K-NN technique has the advantage of being an analytically tractable classifier for an IDS. The Euclidean distance [69] is given as.

$$d(X, Z) = \sqrt{\sum_{I=1}^{N}(Zi - Xi)^2} \quad (6)$$

In this paper, the KNN was used for the classification of the ten components attributes selected by PCA. The calibration plot of KNN is shown in Fig. 3.

3.6. Support vector Machine

The SVM is a prevalent, general and useful classification algorithm that can handle binary classification issues [70]. In the SVM classification algorithm, a hyper-plane is used to distinguish the positive class variable from the negative class variable using the structural risk minimization value [71]. SVM gives good generalization power, robust against local minima, and is represented by small parameters [72]. The feature importance selected by the SVM in this paper is shown in Fig. 4.

3.7. Quadratic discriminant analysis

Quadratic discriminant analysis (QDA) is the next-generation classifier in the family of discriminant analysis. QDA gives better results analysis than LDA [73]. It divides observation using the idea of a quadratic function [74]. In this paper, the QDA was used for the classification of the ten components attributes selected by PCA. The calibration plot of QDA in the analysis is shown in Fig. 5.

4. €

4.1. Naïve Bayes

NB is a simple extremely scalable [75] classifier that is grounded on the Bayes Theorem [76]. NB is used to predict the probability of a class belonging to either normal or attack classes [77]. It operates easily in training and classification phases. NB assumes that all attributes in the vector are similarly independent and important [78]. In this paper, the NB was used for the classification of the ten components attributes

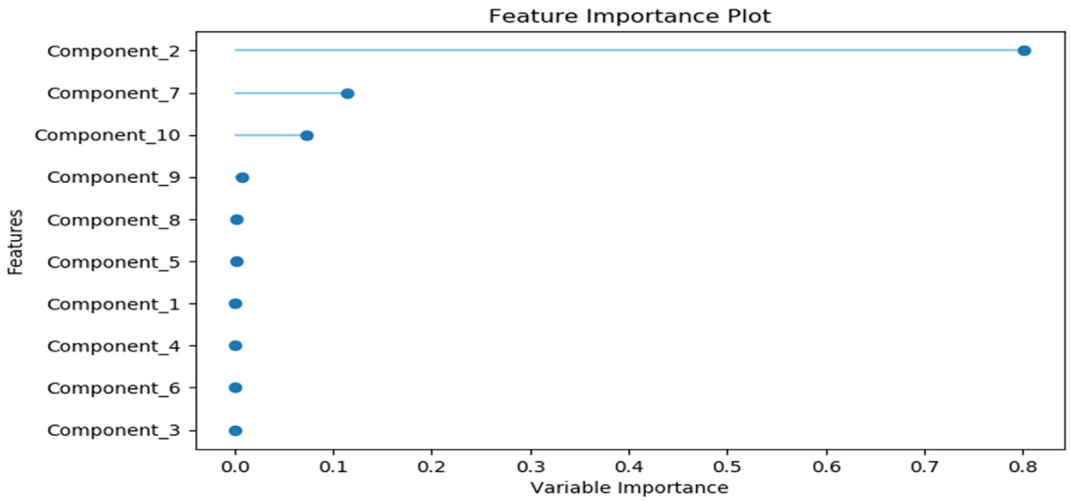

Fig. 2 Xgboost feature importance.

selected by PCA. The calibration plot of NB in the experimental analysis is shown in Fig. 6.

5. Results and discussion

The experimental analysis findings were reported and discussed in this section. This paper adopted the UNSW-NB15. The UNSW-NB15 was formed by using IXIA perfect tool to extract a mixture of contemporary attacks and normal activities of network traffic [47]. The attacks of UNSW-NB15 were grouped into nine different attack types. They are Analysis, Backdoor, DoS, Exploit, Generic, Reconnaissance, Fuzzers, Shellcode, and Worm. The dataset was split into two, where 75% of the dataset was used for training the model, and 25% was used for testing the model.

5.1. Performance measure

The performance of the model was evaluated in terms of the accuracy, area under the curve, recall, precision, F1, kappa, and Mathew Correlation Coefficient. The experimental results of our findings of the proposed models are presented in Table 3.

To corroborate the results of our findings, the decision boundary of each of the models was also generated. The decision boundary gives meaningful insight into how each of the models has studied the task. The decision boundary of XgBoost gives insight into how XgBoost studies the classification task, as demonstrated in Fig. 7.

The decision boundary of KNN gives insight into how KNN studies the classification task, as demonstrated in Fig. 8.

The decision boundary of SVM gives insight into how SVM studies the classification task, as demonstrated in Fig. 9.

The decision boundary of QDA gives insight into how QDA studies the classification task, as demonstrated in Fig. 10.

The decision boundary of NB gives insight into how NB studies the classification task as demonstrated in Fig. 11.

In addition, the results of our findings were compared against recent studies that applied IDS to IoT. Rathore and Park [79] proposed a semi-supervised machine learning technique for IoT. The experimental analysis of their work was done on NSL-KDD. The dataset used in their work suffers from numerous issues, according to the authors in [80]. We think that this dataset should not be adopted for IoT as it was obtained from a conventional network [32]. The majority of datasets utilized in previous papers lack real-world features. This is why the majority of anomalous intrusion detection systems in IoT are unsuitable for use in a production environment. Additionally, they are incapable of adjusting to the continual changes in network architecture. This necessitates us to implement IDS that satisfied IoT protocol requirements like the Wireless low-power personal area network. Therefore, IDS that is made for the IoT environment should work under high-speed settings, big data capacity, low power consumption, and processing. The authors in [81] also adopted the NSL-KDD dataset. This same dataset limitation was the major problem in their work. The work in [82] suffers in terms of accuracy and dataset used for the experimental analysis,

Table 2 Cat Boost Algorithm.

1:	Input: $\{Yi, Zi\}_{i=1}^{n}$, J
2:	Output: $\sigma \leq [1, u]$
3:	$M_j \leq 0$ for L = 1...u;
4:	For r $\leq$ 1 to J do
5:	For l $\leq$ 1 to u do
6:	$T_l \leq s_l - M_{(j)} - 1\ (Y_j)$;
7:	For L $\leq$ 1 to u do
8:	λ M $\leq$ Learning model ((x, t), σ (1) $\leq$ l);
9:	$M_l \leq M_l + \lambda$ M;
10:	Return M_u

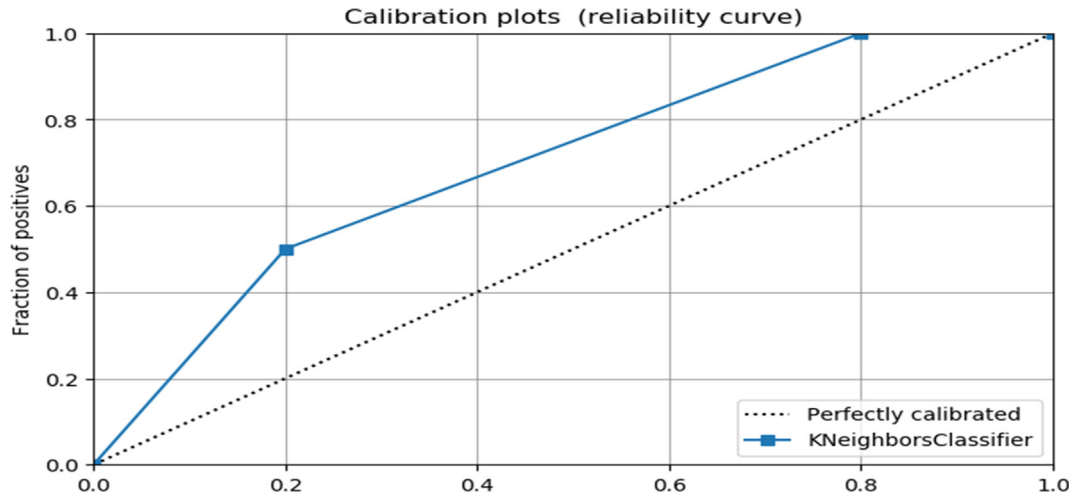

Fig. 3 Calibration plot of KNN.

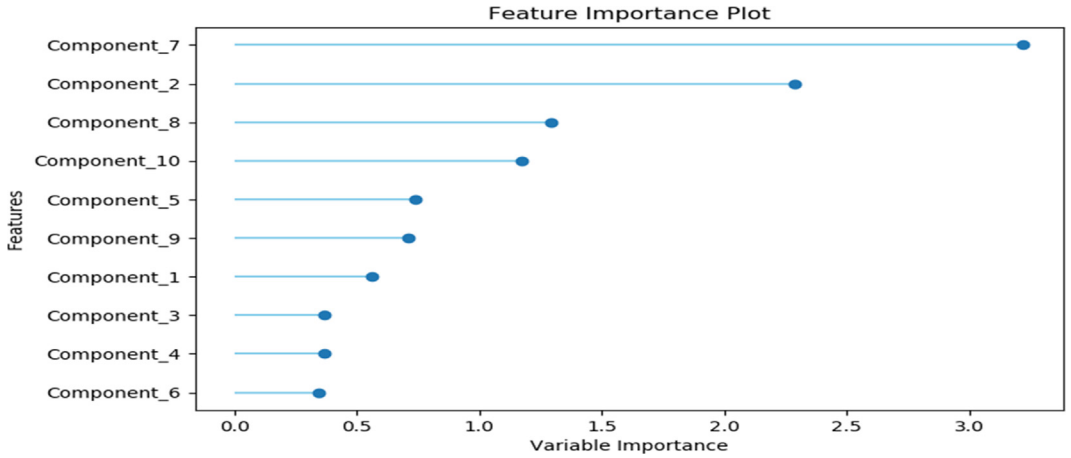

Fig. 4 Feature Importance Selected by SVM.

which does not reflect contemporary attacks and is not a good fit for the IoT environment. The study in [34] also adopted a dataset that is of a lower standard to NSLKDD. The main issues in their work were the dataset employed for analysis and low accuracy. The authors in [84]-[88] evaluated the performances of their models in terms of accuracy, precision, and F1 without considering the MCC. However, we extend the performance metrics in our proposed study by introducing

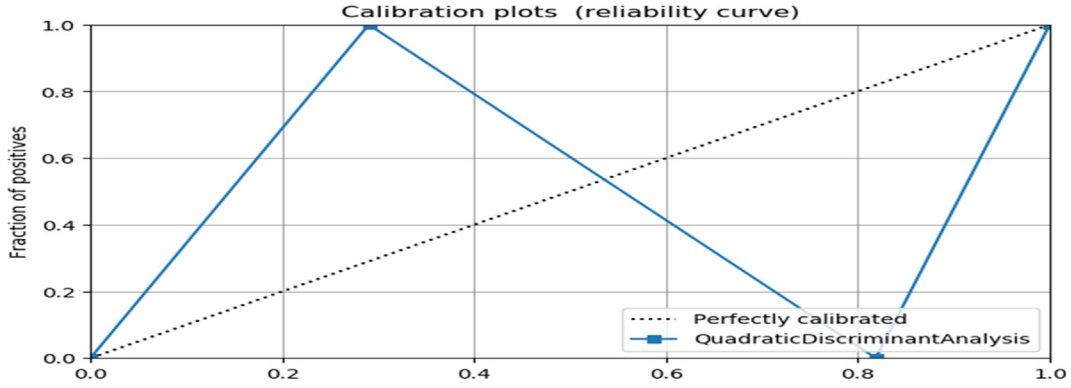

Fig. 5 Calibration plot of QDA.

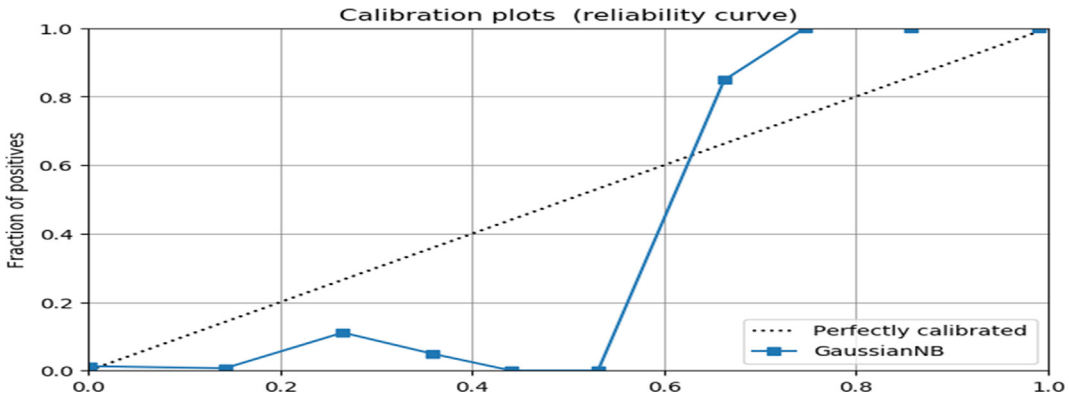

Fig. 6 Calibration plot of NB.

Table 3 Performance measures of the proposed model.

Classifiers	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC	Training Time
PCA-XGBoost	99.99	1.00	99.99	1.00	1.00	99.97	99.97	0.7094
PCA-Cat Boost	99.99	1.00	99.99	1.00	99.99	99.97	99.97	18.090
PCA-KNN	99.98	1.00	99.98	1.00	99.99	99.96	99.96	0.0930
PCA-SVM	99.98	0.00	99.98	1.00	99.99	99.96	99.96	0.0322
PCA-QDA	99.97	1.00	99.97	99.99	99.98	99.94	99.94	0.1142
PCA-NB	97.14	99.77	99.2	96.72	97.94	93.28	93.41	0.0102

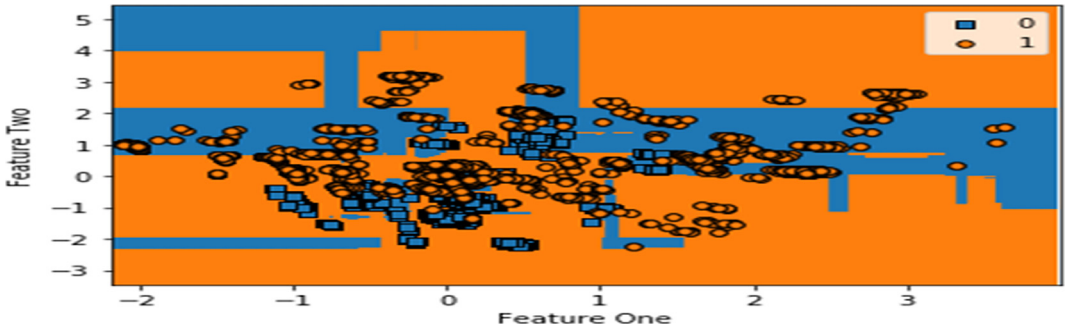

Fig. 7 Decision Boundary of XgBoost.

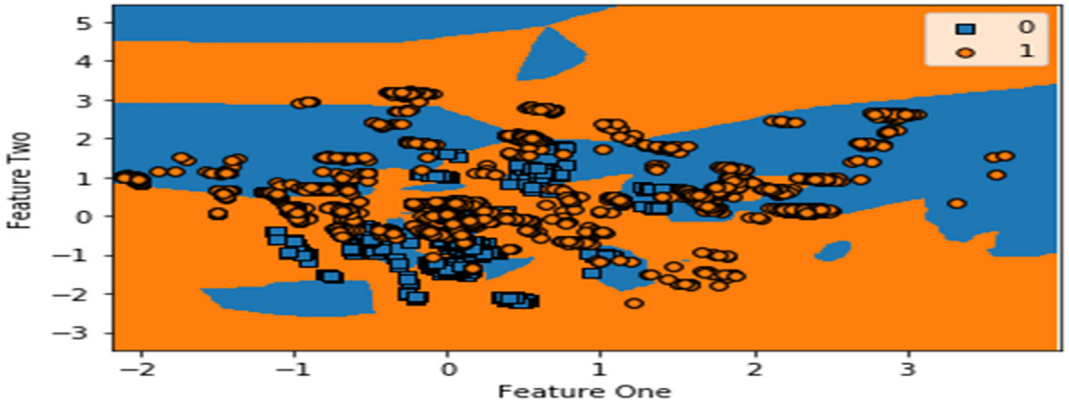

Fig. 8 Decision Boundary of KNN.

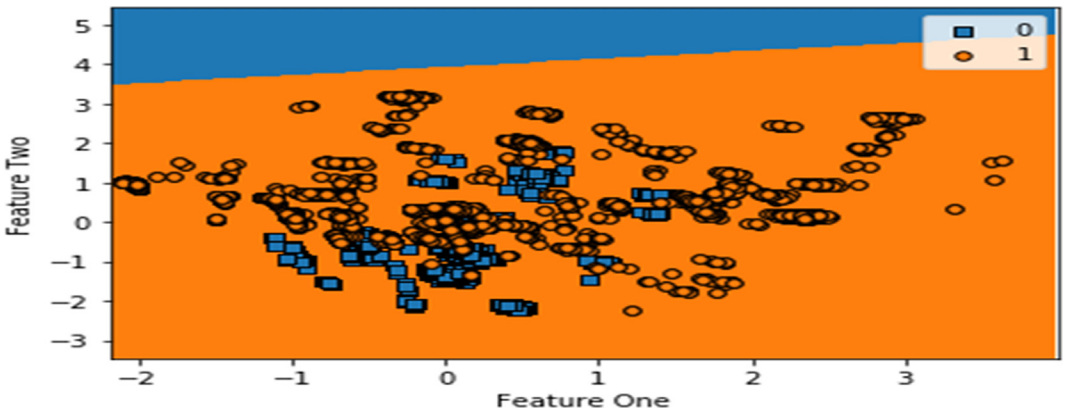

Fig. 9 Decision Boundary of SVM.

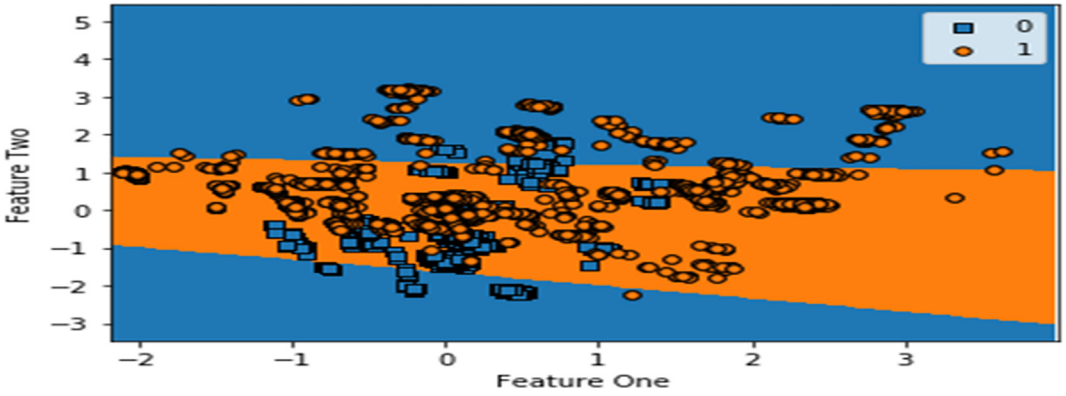

Fig. 10 Decision Boundary of QDA.

the MCC. Our Proposed methods outperformed all reported studies, as shown in Table 4 in terms of validation dataset, accuracy, precision, F1 score, and MCC.

The accuracy of our proposed method PCA-XgBoost gave the highest accuracy as revealed in Table 3 and Fig. 12 in terms of accuracy, precision, F1, and MCC than all other proposed methods. The PCA-Cat Boost also outperformed other proposed methods with an accuracy of 99.99, precision of 1, F1 of 99.99, and MCC of 99.97.

From Fig. 12, we observed that the PCA-KNN also gave an accuracy of 99.98, precision of 1, F1 of 99.99, and MCC of 99.96. Our proposed PCA-SVM gave an accuracy of 99.98, precision of 0, F1 of 1, and MCC of 99.96. The proposed PCA-QDA gave an accuracy of 99.97, precision of 99.99, F1 of 99.98, and MCC of 99.94. Lastly, the proposed PCA-NB gave an accuracy of 97.14, precision of 96.72, F1 of 97.94, and MCC of 93.41. In the experimental testbed situation, we discover that our proposed models surpass the previ-

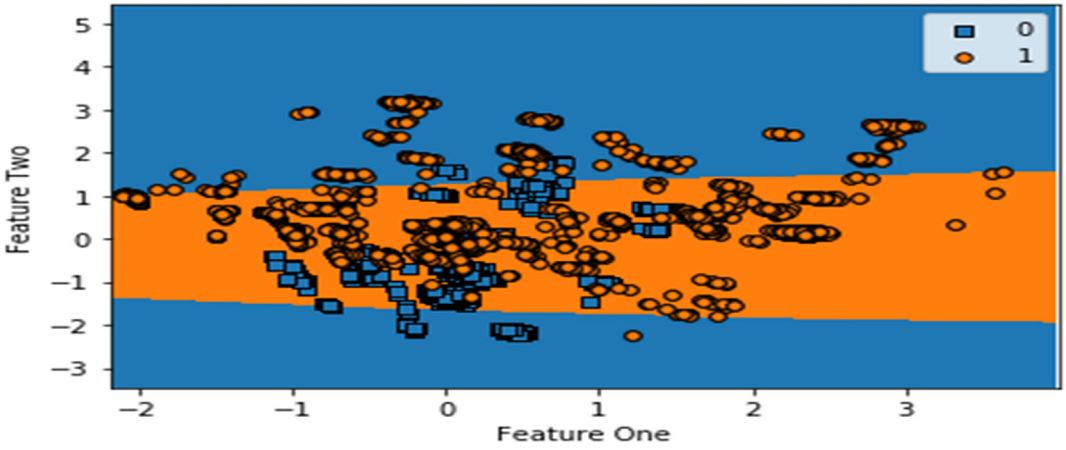

Fig. 11 Decision Boundary of NB.

Table 4 Comparison with past studies.

Authors	Security Threat	Validation Dataset Strategy	Accuracy	Precision	F1 Score	MCC
[79]	Network	NSL-KDD	86.53	–	–	–
[83]	Network	Network Traffic Data	99.49	–	–	–
[81]	Network	NSL-KDD	94.27	92.18	92.29	84.44
[82]	Network	NSL-KDD	91.39	–	–	–
[34]	Network	KDD-Cup 99	96.8	–	–	–
[84]	Network	NSL-KDD	98	97	97	–
[85]	Network	KDD-Cup 99	94	98	96	–
[86]	Network	UNSW-NB15	98.6	1	1	–
[87]	Network	DDoS	98.7	95	97	–
[88]	Network	CICIDS	98.3	90.0	91.67	–
Proposed XgBoost	Network	UNSWNB-15	99.99	1.00	1.00	99.97
Proposed CatBoost	Network	UNSW-NB15	99.99	1.00	99.99	99.97
Proposed KNN	Network	UNSW-NB15	99.98	1.00	99.99	99.96
Proposed SVM	Network	UNSW-NB15	99.98	1.00	99.99	99.96
Proposed QDA	Network	UNSW-NB15	99.97	99.99	99.98	99.94
Proposed NB	Network	UNSW-NB15	97.14	96.72	97.94	93.41

ous detection technique in terms of accuracy and MCC. It is critical to highlight that our detection system was previously trained before these experiments, and so the training time remains constant. These results demonstrate that our suggested IDS perform well on simulated network traffic. As a result, we can conclude that our suggested system is capable of reliably detecting security assaults in a variety of network attack scenarios.

5.2. Threats to validity

This section looks at potential challenges to the validity of the verification results acquired in this research.

5.2.1. Internal validity

Internal validity is the degree to which reported results accurately reflect the reality in the population under investigation and are not the product of methodological flaws. There are two key aspects to consider here.

5.2.2. Instrumentation

This refers to discrepancies caused by changes in an instrument's calibration, as well as changes in the scorers, observers, or probably the device itself. The accuracy, validation dataset, precision, MCC, and F1 metrics used in the validation are all well-known methodologies. As a result, there have been no modifications that could have caused the evaluation results to be incorrect.

5.2.3. Selection

Any factor besides the system that causes posttest disparities is referred to as a selection threat. Therefore, the situation in which the feature scaling is not performed and the data is not on the same scale can be a factor in this work.

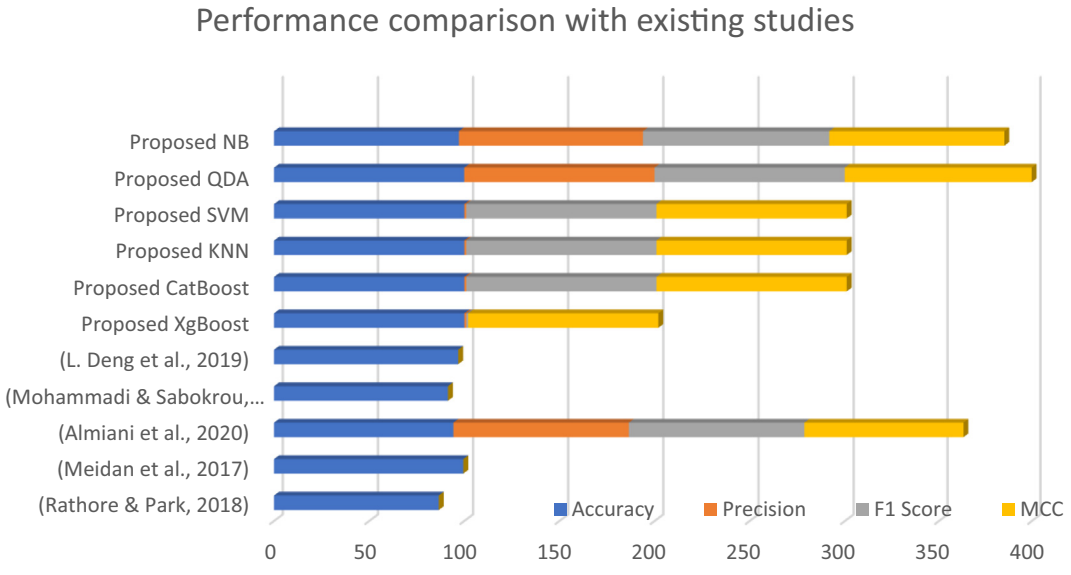

Fig. 12 Performance comparison with the state-of-the-art.

5.3. Construct validity

The degree to which the instrument 'interacts' in a way that is consistent with conceptual assumptions, and how effectively the instrument's scores are reflective of the complex framework. This threat stems from the question of if the experiment accurately replicates real-world occurrences to be examined. The evaluation criteria in terms of accuracy are very high indicating the proposed model is consistent.

5.4. External validity

This has to do with our ability to apply the findings of this research to real-world situations. This threat raises the question, "Can this effect be generalized to various populations, situations, treatment, and measurement attributes?".

On the UNSW-NB15 data, the proposed ML technique for IoT network threat detection was applied and confirmed. The findings are consistent with what has been found in the literature. Validation will be done in an industry context or on the IoT botnet dataset in the future.

6. Conclusion and future work

We examined the viability of deploying machine-learning-based intrusion detection in resource-constrained IoT environments in this paper. To that aim, we built an intelligent IDS capable of detecting abnormal behavior on insecure IoT networks by deftly combining feature dimensionality reduction and machine learning methods. We evaluated our scheme's performance using the UNSW-NB15 dataset to determine the optimal approach for machine learning-based IDS. Security concerns have become a major roadblock to the development of the IoT. Security detection tasks could be handled by machine learning-based IDS. The PCA algorithm was used for dimensionality reduction to select ten components. The model was evaluated on a recent dataset UNSW-NB15 that supports contemporary attacks and is very appropriate for IoT applications. Also, communication overhead is reduced in the proposed model, and there is no need for a foreign key as required in encryption methods for IoT network security. Our results revealed that the suggested approach achieves higher F1 scores, indicating a stronger overall detection performance. Based on the experimental results from network simulations and testbed implementations, we can infer that using machine learning techniques for successful anomaly detection in the IoT environment is both realistic and practicable. The comparison of the proposed PCA-XgBoost, PCA-Cat Boost, PCA-KNN, PCA-SVM, PCA-QDA, and PCA-NB with existing studies demonstrates outstanding accuracy and can address the issue of labeled data in IoT applications. The experimental findings of our work were superior to the state-of-the-art in terms of validation dataset, precision, F1, MCC, and accuracy of the two of our proposed models, attaining 99.99%. The architecture can be organized with smart cities, smart homes, and healthcare devices as a unit that detects attacks in the IoT settings. The future work will be to adopt an ensemble model with a novel dataset suitable for the IoT environment. In addition, this approach can be enhanced in the future by incorporating deep learning models. More specifically, the BoT-IoT dataset will be utilized for experimental analysis and compared to the UNSWNB-15 with a deep learning model for network traffic classification.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This paper is partially funded by the Research Council of Norway (RCN) in the INTPART program under the project "Reinforcing Competence in Cybersecurity of Critical Infrastructures: A Norway-US Partnership (RECYCIN)" with the project number #309911.

References

[1] B.B. Zarpelão, R.S. Miani, C.T. Kawakani, S.C. de Alvarenga, A survey of intrusion detection in Internet of Things, J. Netw. Comput. Appl. 84 (2017) 25–37, https://doi.org/10.1016/j.jnca.2017.02.009.

[2] D. Miorandi, S. Sicari, F. De Pellegrini, I. Chlamtac, Internet of things: Vision, applications and research challenges, Ad Hoc Networks 10 (7) (2012) 1497–1516, https://doi.org/10.1016/j.adhoc.2012.02.016.

[3] A.B. Feroz Khan, G. Anandharaj, A Multi-layer Security approach for DDoS detection in Internet of Things, Int. J. Intell. Unmanned Syst. 9 (3) (2020) 178–191, https://doi.org/10.1108/IJIUS-06-2019-0029.

[4] "Cisco Delivers Vision of Fog Computing to Accelerate Value from Billions of Connected Devices | The Network." https://newsroom.cisco.com/press-release-content?type=webcontent&articleId=1334100 (accessed Nov. 30, 2020).

[5] S. Sicari, A. Rizzardi, L.A. Grieco, A. Coen-Porisini, Security, privacy and trust in Internet of things: The road ahead, Comput. Networks 76 (2015) 146–164, https://doi.org/10.1016/j.comnet.2014.11.008.

[6] J. Jin, J. Gubbi, S. Marusic, M. Palaniswami, An information framework for creating a smart city through internet of things, IEEE Internet Things J. 1 (2) (2014) 112–121, https://doi.org/10.1109/JIOT.2013.2296516.

[7] D. Singh, G. Tripathi, A.J. Jara, "A survey of Internet-of-Things: Future vision, architecture, challenges and services", *2014 IEEE World Forum Internet Things*, WF-IoT 2014 (2014) 287–292, https://doi.org/10.1109/WF-IoT.2014.6803174.

[8] C. Perera, C.H. Liu, S. Jayawardena, M. Chen, A Survey on Internet of Things from Industrial Market Perspective, IEEE Access 2 (2015) 1660–1679, https://doi.org/10.1109/ACCESS.2015.2389854.

[9] H. A. Abdul-Ghani and D. Konstantas, "A comprehensive study of security and privacy guidelines, threats, and countermeasures: An IoT perspective," *J. Sens. Actuator Networks*, vol. 8, no. 2, 2019, doi: 10.3390/jsan8020022.

[10] V. Adat and B. B. Gupta, "Security in Internet of Things: issues, challenges, taxonomy, and architecture," *Telecommun. Syst.*, vol. 67, no. 3, pp. 423–441, 2018, doi: 10.1007/s11235-017-0345-9.

[11] E. Anthi, L. Williams, P. Burnap, Pulse: An adaptive intrusion detection for the internet of things, IET Conf. Publ. 2018 (CP740) (2018) 1–4, https://doi.org/10.1049/cp.2018.0035.

[12] S. Cirani, G. Ferrari, L. Veltri, Enforcing security mechanisms in the IP-based internet of things: An algorithmic overview, Algorithms 6 (2) (2013) 197–226, https://doi.org/10.3390/a6020197.

[13] C. Thirumalai, S. Mohan, G. Srivastava, An efficient public key secure scheme for cloud and IoT security, Comput. Commun. 150 (2020) 634–643, https://doi.org/10.1016/j.comcom.2019.12.015.

[14] A. Riahi Sfar, E. Natalizio, Y. Challal, Z. Chtourou, A roadmap for security challenges in the internet of things, Digit Commun Netw 4 (2) (2018) 118–137.

[15] M. Zolanvari, M.A. Teixeira, L. Gupta, K.M. Khan, R. Jain, Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things, IEEE Internet Things J. 6 (4) (2019) 6822–6834, https://doi.org/10.1109/JIOT.2019.2912022.

[16] A. Alrawais, A. Alhothaily, C. Hu, X. Cheng, Fog Computing for the Internet of Things: Security and Privacy Issues, IEEE Internet Comput. 21 (2) (2017) 34–42, https://doi.org/10.1109/MIC.2017.37.

[17] Y. K. Saheed, "Performance Improvement of Intrusion Detection System for Detecting Attacks on Internet of Things and Edge of Things," in *Artificial Intelligence for Cloud and Edge Computing. Internet of Things (Technology, Communications and Computing)*, S. Misra, T. K. A., V. Piuri, and L. Garg, Eds. Springer, Cham, 2022, pp. 321–339.

[18] A.P. Kelton, J.P. Papa, C.O. Lisboa, R. Munoz, V.H.C. De, Internet of Things : A survey on machine learning-based intrusion detection approaches, Comput. Networks 151 (2019) 147–157, https://doi.org/10.1016/j.comnet.2019.01.023.

[19] W. Wu, H. Zhang, S. Pirbhulal, S.C. Mukhopadhyay, Y.T. Zhang, Assessment of Biofeedback Training for Emotion Management Through Wearable Textile Physiological Monitoring System, IEEE Sens. J. 15 (12) (2015) 7087–7095, https://doi.org/10.1109/JSEN.2015.2470638.

[20] D. Pasini, S. Mastrolembo Ventura, S. Rinaldi, P. Bellagente, A. Flammini, and A. L. C. Ciribini, "Exploiting internet of things and building information modeling framework for management of cognitive buildings," *IEEE 2nd Int. Smart Cities Conf. Improv. Citizens Qual. Life, ISC2 2016 - Proc.*, vol. 40545387, no. 40545387, 2016, doi: 10.1109/ISC2.2016.7580817.

[21] W. Wu, S. Pirbhulal, H. Zhang, S.C. Mukhopadhyay, Quantitative Assessment for Self-Tracking of Acute Stress Based on Triangulation Principle in a Wearable Sensor System, IEEE J. Biomed. Heal. Informatics 23 (2) (2019) 703–713, https://doi.org/10.1109/JBHI.2018.2832069.

[22] E. Kabir, J. Hu, H. Wang, G. Zhuo, A novel statistical technique for intrusion detection systems, Futur. Gener. Comput. Syst. 79 (2018) 303–318, https://doi.org/10.1016/j.future.2017.01.029.

[23] M. Ahmed, A. Naser Mahmood, J. Hu, A survey of network anomaly detection techniques, J. Netw. Comput. Appl. 60 (2016) 19–31, https://doi.org/10.1016/j.jnca.2015.11.016.

[24] A.A. Diro, N. Chilamkurti, Distributed Attack Detection Scheme using Deep Learning Approach for Internet of Things, Futur. Gener. Comput. Syst. 82 (2018) 761–768, https://doi.org/10.1016/j.future.2017.08.043.

[25] S.R. Nabavi, S.M. Mousavi, "A Novel Cluster-based Key Management Scheme to Improve Scalability in Wireless Sensor Networks" 16 (7) (2016) 150–156.

[26] S.D. Babar, P.N. Mahalle, A Hash Key-Based Key Management Mechanism for Cluster-Based Wireless Sensor Network, J. Cyber Secur. Mobil. 5 (2017) 73–88, https://doi.org/10.13052/2245-1439.524.

[27] P. Kasinathan, C. Pastrone, M.A. Spirito, M. Vinkovits, Denial-of-Service detection in 6LoWPAN based Internet of Things, Int. Conf. Wirel. Mob. Comput. Netw. Commun. (2013) 600–607, https://doi.org/10.1109/WiMOB.2013.6673419.

[28] J.H. Anajemba, Y. Tang, C. Iwendi, A. Ohwoekevwo, G. Srivastava, O. Jo, Realizing efficient security and privacy in IoT networks, Sensors (Switzerland) 20 (9) (2020) 1–24, https://doi.org/10.3390/s20092609.

[29] A.B. Feroz Khan, G. Anandharaj, A cognitive key management technique for energy efficiency and scalability in securing the sensor nodes in the IoT environment: CKMT, SN Appl. Sci. 1 (12) (2019), https://doi.org/10.1007/s42452-019-1628-4.

[30] P. Vijayakumar, M. Azees, V. Chang, J. Deborah, B. Balusamy, Computationally efficient privacy preserving authentication and key distribution techniques for vehicular ad hoc networks, Cluster Comput. 20 (3) (2017) 2439–2450, https://doi.org/10.1007/s10586-017-0848-x.

[31] Y.K. Saheed, M.O. Arowolo, Efficient Cyber Attack Detection on the Internet of Medical Things-Smart Environment Based on Deep Recurrent Neural Network and Machine Learning Algorithms, IEEE Access 9 (2021) 161546–161554, https://doi.org/10.1109/ACCESS.2021.3128837.

[32] A. Khraisat, I. Gondal, P. Vamplew, J. Kamruzzaman, and A. Alazab, "A novel ensemble of hybrid intrusion detection system for detecting internet of things attacks," *Electron.*, vol. 8, no. 11, 2019, doi: 10.3390/electronics8111210.

[33] J. John, M.S. Varkey, M. Selvi, Security attacks in s-wbans on iot based healthcare applications, Int. J. Innov. Technol. Explor. Eng. 9 (1) (2019) 2088–2097, https://doi.org/10.35940/ijitee.A4242.119119.

[34] L. Deng, D. Li, X. Yao, D. Cox, H. Wang, Mobile network intrusion detection for IoT system based on transfer learning algorithm, Cluster Comput. 22 (2019) 9889–9904, https://doi.org/10.1007/s10586-018-1847-2.

[35] A. Adnan, A. Muhammed, A.A.A. Ghani, A. Abdullah, F. Hakim, An intrusion detection system for the internet of things based on machine learning: Review and challenges, Symmetry (Basel) 13 (6) (2021) 1–13, https://doi.org/10.3390/sym13061011.

[36] E. Hodo *et al.*, "Threat analysis of IoT networks using artificial neural network intrusion detection system," *2016 Int. Symp. Networks, Comput. Commun. ISNCC 2016*, pp. 4–9, 2016, doi: 10.1109/ISNCC.2016.7746067.

[37] Q. Niyaz, W. Sun, A.Y. Javaid, M. Alam, "A deep learning approach for network intrusion detection system", *EAI Int*, Conf. Bio-inspired Inf. Commun. Technol. (2015), https://doi.org/10.4108/eai.3-12-2015.2262516.

[38] H. Bostani, M. Sheikhan, Hybrid of anomaly-based and specification-based IDS for Internet of Things using unsupervised OPF based on MapReduce approach, Comput. Commun. 98 (2017) 52–71, https://doi.org/10.1016/j.comcom.2016.12.001.

[39] H.H. Pajouh, R. Javidan, R. Khayami, A. Dehghantanha, K.-K. Choo, A Two-Layer Dimension Reduction and Two-Tier Classification Model for Anomaly-Based Intrusion Detection in IoT Backbone Networks, IEEE Trans. Emerg. Top. Comput. 7 (2) (2019) 314–323, https://doi.org/10.1109/TETC.2016.2633228.

[40] R. Kozik, M. Choraś, M. Ficco, F. Palmieri, A scalable distributed machine learning approach for attack detection in edge computing environments, J. Parallel Distrib. Comput. 119 (2018) 18–26, https://doi.org/10.1016/j.jpdc.2018.03.006.

[41] M. Tsikala Vafea, E. Atalla, J. Georgakas, F. Shehadeh, E.K. Mylona, M. Kalligeros, E. Mylonakis, Emerging Technologies for Use in the Study, Diagnosis, and Treatment of Patients with COVID-19, Cell. Mol. Bioeng. 13 (4) (2020) 249–257, https://doi.org/10.1007/s12195-020-00629-w.

[42] M. Otoom, N. Otoum, M.A. Alzubaidi, Y. Etoom, R. Banihani, Biomedical Signal Processing and Control An IoT-based framework for early identification and monitoring of COVID-19 cases, Biomed. Signal Process. Control 62 (2020) 102149, https://doi.org/10.1016/j.bspc.2020.102149.

[43] S. Kumar, R.D. Raut, B.E. Narkhede, A proposed collaborative framework by using artificial intelligence-internet of things (AI-IoT) in COVID-19 pandemic situation for healthcare workers, Int. J. Healthc. Manag. 13 (4) (2020) 337–345, https://doi.org/10.1080/20479700.2020.1810453.

[44] Y. Feng, J. Zhong, C.X. Ye, Z.F. Wu, "Clustering based on self-organizing ant colony networks with application to intrusion detection", *Proc. - ISDA 2006 Sixth Int*, Conf. Intell. Syst. Des. Appl. 2 (2006) 1077–1080, https://doi.org/10.1109/ISDA.2006.253761.

[45] Y.W. Chen, J.P. Sheu, Y.C. Kuo, N. Van Cuong, "Design and implementation of IoT DDoS attacks detection system based on machine learning", *2020 Eur*, Conf. Networks Commun. EuCNC 2020 (2020) 122–127, https://doi.org/10.1109/EuCNC48522.2020.9200909.

[46] M. Ahmad, Q. Riaz, M. Zeeshan, H. Tahir, S.A. Haider, M.S. Khan, Intrusion detection in internet of things using supervised machine learning based on application and transport layer features using UNSW-NB15 data-set, Eurasip J. Wirel. Commun. Netw. 1 (2021) 2021, https://doi.org/10.1186/s13638-021-01893-8.

[47] N. Moustafa and J. Slay, "The significant features of the UNSW-NB15 and the KDD99 data sets for Network Intrusion Detection Systems," *Proc. - 2015 4th Int. Work. Build. Anal. Datasets Gather. Exp. Returns Secur. BADGERS 2015*, pp. 25–31, 2017, doi: 10.1109/BADGERS.2015.14.

[48] E.A. Felix, S.P. Lee, Systematic literature review of preprocessing techniques for imbalanced data, IET Softw. 13 (6) (2019) 479–496, https://doi.org/10.1049/iet-sen.2018.5193.

[49] Y. Zhou, G. Cheng, S. Jiang, M. Dai, Building an efficient intrusion detection system based on feature selection and ensemble classifier, Comput. Networks 174 (October) (2019) 2020, https://doi.org/10.1016/j.comnet.2020.107247.

[50] S. Jain, S. Shukla, R. Wadhvani, Dynamic selection of normalization techniques using data complexity measures, Expert Syst. Appl. 106 (2018) 252–262, https://doi.org/10.1016/j.eswa.2018.04.008.

[51] S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014.

[52] H. Alazzam, A. Sharieh, K.E. Sabri, A feature selection algorithm for intrusion detection system based on Pigeon Inspired Optimizer, Expert Syst. Appl. 148 (2020) 113249, https://doi.org/10.1016/j.eswa.2020.113249.

[53] S. Maza, M. Touahria, Feature selection for intrusion detection using new multi-objective estimation of distribution algorithms, Appl. Intell. 49 (12) (2019) 4237–4257, https://doi.org/10.1007/s10489-019-01503-7.

[54] F.H. Almasoudy, W.L. Al-Yaseen, A.K. Idrees, Differential Evolution Wrapper Feature Selection for Intrusion Detection System, Procedia Comput. Sci. 167 (2019) (2020) 1230–1239, https://doi.org/10.1016/j.procs.2020.03.438.

[55] Y.K. Saheed, F.E. Hamza-Usman, Feature Selection with IG-R for Improving Performance of Intrusion Detection System, Int. J. Commun. Networks Inf. Secur 12 (3) (2020) 338–344.

[56] A. Yulianto, P. Sukarno, and N. A. Suwastika, "Improving AdaBoost-based Intrusion Detection System (IDS) Performance on CIC IDS 2017 Dataset," *J. Phys. Conf. Ser.*, vol. 1192, no. 1, 2019, doi: 10.1088/1742-6596/1192/1/012018.

[57] R. Abdulhammed, H. Musafer, A. Alessa, M. Faezipour, A. Abuzneid, Features Dimensionality Reduction Approaches for Machine Learning Based Network Intrusion Detection, Electronics 8 (3) (2019) 322, https://doi.org/10.3390/electronics8030322.

[58] J. Gao, S. Chai, B. Zhang, and Y. Xia, "Research on network intrusion detection based on incremental extreme learning machine and adaptive principal component analysis," *Energies*, vol. 12, no. 7, 2019, doi: 10.3390/en12071223.

[59] S. Bhattacharya *et al.*, "A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU," *Electron.*, vol. 9, no. 2, 2020, doi: 10.3390/electronics9020219.

[60] S. Velliangiri, "A hybrid BGWO with KPCA for intrusion detection," J. Exp. Theor. Artif. Intell., vol. 32, no. 1, pp. 165–180, 2020, doi: 10.1080/0952813X.2019.1647558.

[61] D. Gonzalez-Cuautle et al., "Synthetic minority oversampling technique for optimizing classification tasks in botnet and intrusion-detection-system datasets," Appl. Sci., vol. 10, no. 3, 2020, doi: 10.3390/app10030794.

[62] C. Hu, J. Yan, and C. Wang, "Advanced Cyber-Physical Attack Classification with Extreme Gradient Boosting for Smart Transmission Grids," *IEEE Power Energy Soc. Gen. Meet.*, vol. 2019-Augus, 2019, doi: 10.1109/PESGM40551.2019.8973679.

[63] A. Husain, A. Salem, C. Jim, and G. Dimitoglou, "Development of an Efficient Network Intrusion Detection Model Using Extreme Gradient Boosting (XGBoost) on the UNSW-NB15 Dataset," *2019 IEEE 19th Int. Symp. Signal Process. Inf. Technol. ISSPIT 2019*, 2019, doi: 10.1109/ISSPIT47144.2019.9001867.

[64] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: Gradient boosting with categorical features support," *arXiv*, pp. 1–7, 2018.

[65] T. Al-hadhrami and F. Mohammed, *Advances on Smart and Soft Computing*. 2020.

[66] G. Kavitha, N.M. Elango, An approach to feature selection in intrusion detection systems using machine learning algorithms, Int. J. e-Collaboration 16 (4) (2020) 48–58, https://doi.org/10.4018/IJeC.2020100104.

[67] G. Serpen, E. Aghaei, Host-based misuse intrusion detection using PCA feature extraction and kNN classification algorithms, Intell. Data Anal. 22 (5) (2018) 1101–1114, https://doi.org/10.3233/IDA-173493.

[68] N. Moustafa, J. Slay, "A hybrid feature selection for network intrusion detection systems: Central points and association rules" arXiv (2017) 5–13, https://doi.org/10.4225/75/57a84d4fbefbb.

[69] A.A. Salih, M.B. Abdulrazaq, Combining Best Features Selection Using Three Classifiers in Intrusion Detection System, 2019 Int Conf. Adv. Sci. Eng. ICOASE 2019 (2019) 94–99, https://doi.org/10.1109/ICOASE.2019.8723671.

[70] W. Wang, X. Du, N. Wang, Building a Cloud IDS Using an Efficient Feature Selection Method and SVM, IEEE Access 7 (2019) 1345–1354, https://doi.org/10.1109/ACCESS.2018.2883142.

[71] M. Al-Qatf, Y. Lasheng, M. Al-Habib, K. Al-Sabahi, Deep Learning Approach Combining Sparse Autoencoder with SVM for Network Intrusion Detection, IEEE Access 6 (2018) 52843–52856, https://doi.org/10.1109/ACCESS.2018.2869577.

[72] W. Feng, J. Sun, L. Zhang, C. Cao, Q. Yang, A support vector machine based naive Bayes algorithm for spam filtering, 2016 IEEE 35th Int. Perform. Comput. Commun. Conf. IPCCC 2016 (2017), https://doi.org/10.1109/PCCC.2016.7820655.

[73] P. P. S. Saputra, F. D. Murdianto, R. Firmansyah, and K. Widarsono, "Combination of Quadratic Discriminant Analysis and Daubechis Wavelet for Classification Level of Misalignment on Induction Motor," *Proceeding - 2019 Int. Symp. Electron. Smart Devices, ISESD 2019*, pp. 1–5, 2019, doi: 10.1109/ISESD.2019.8909431.

[74] Y. Saheed, O. Longe, U. A. Baba, S. Rakshit, and N. R. Vajjhala, "An Ensemble Learning Approach for Software Defect Prediction in Developing Quality Software Product.," in *Advances in Computing and Data Sciences.*, M. Singh, V. Tyagi, P. K. Gupta, J. Flusser, T. Ören, and V. R. Sonawane, Eds. Springer, Cham, 2021.

[75] M.O. Mughal, S. Kim, S. Member, "Signal Classification and Jamming Detection in Wide-band Radios Using Na¨ıve, Bayes Classifier" 14 (8) (2018) 8–11, https://doi.org/10.1109/LCOMM.2018.2830769.

[76] S.M. Kasongo, Y. Sun, A deep learning method with filter based feature engineering for wireless intrusion detection system, IEEE Access 7 (2019) 38597–38607, https://doi.org/10.1109/ACCESS.2019.2905633.

[77] J. Manhas, S. Kotwal, Implementation of Intrusion Detection System for Internet of Things Using Machine Learning Techniques, Multimedia Security. Algorithms Intelligent Systems (2021), https://doi.org/10.1007/978-981-15-8711-5_11.

[78] L.I. Li, W. Jiang, X. Li, K.L. Moser, Z. Guo, L. Du, Q. Wang, E.J. Topol, Q. Wang, S. Rao, A robust hybrid between genetic algorithm and support vector machine for extracting an optimal feature gene subset, Genomics 85 (1) (2005) 16–23, https://doi.org/10.1016/j.ygeno.2004.09.007.

[79] S. Rathore, J.H. Park, Semi-supervised learning based distributed attack detection framework for IoT, Appl. Soft Comput. J. 72 (2018) 79–89, https://doi.org/10.1016/j.asoc.2018.05.049.

[80] J. Mchugh, Testing Intrusion Detection Systems: A Critique of the 1998 and 1999 DARPA Intrusion Detection System Evaluations as Performed by Lincoln Laboratory, ACM Trans. Inf. Syst. Secur. 3 (4) (2000) 262–294, https://doi.org/10.1145/382912.382923.

[81] M. Almiani, A. AbuGhazleh, A. Al-Rahayfeh, S. Atiewi, A. Razaque, Deep recurrent neural network for IoT intrusion detection system, Simul. Model. Pract. Theory 101 (2020), https://doi.org/10.1016/j.simpat.2019.102031 102031.

[82] B. Mohammadi, M. Sabokrou, "End-to-End Adversarial Learning for Intrusion Detection in Computer Networks" arXiv (2019) 270–273.

[83] Y. Meidan *et al.*, "Detection of Unauthorized IoT Devices Using Machine Learning Techniques," *arXiv*, 2017.

[84] C. Liang, B. Shanmugam, S. Azam, M. Jonkman, F. De Boer, G. Narayansamy, Intrusion Detection System for Internet of Things based on a Machine Learning approach, 2019 Int. Conf. Vis. Towar. Emerg. Trends Commun. Netw. (2019) 1–6.

[85] S. Fenanir, F. Semchedine, A. Baadache, A Machine Learning-Based Lightweight Intrusion Detection System for the Internet of Things, Rev. d'Intelligence Artif. 33 (3) (2019) 203–211.

[86] S.S. Abul Basar, Haoxiang Wang, Hybrid Intrusion Detection System for Internet of Things (IoT), J. ISMAC 2 (4) (2020) 190–199, https://doi.org/10.36548/jismac.2020.4.002.

[87] G. Thamilarasu and S. Chawla, "Towards deep-learning-driven intrusion detection for the internet of things," *Sensors (Switzerland)*, vol. 19, no. 9, 2019, doi: 10.3390/s19091977.

[88] S.U. Jan, S. Ahmed, V. Shakhov, I. Koo, Toward a Lightweight Intrusion Detection System for the Internet of Things, IEEE Access 7 (2019) 42450–42471, https://doi.org/10.1109/ACCESS.2019.2907965.