BIG DATA MINING AND ANALYTICS
ISSN 2096-0654 02/10 pp273–287
Volume 6, Number 3, September 2023
DOI: 10.26599/BDMA.2022.9020032

An Ensemble Learning Based Intrusion Detection Model for Industrial IoT Security

Mouaad Mohy-Eddine, Azidine Guezzaz*, Said Benkirane, Mourade Azrour, and Yousef Farhaoui

Abstract: Industrial Internet of Things (IIoT) represents the expansion of the Internet of Things (IoT) in industrial sectors. It is designed to implicate embedded technologies in manufacturing fields to enhance their operations. However, IIoT involves some security vulnerabilities that are more damaging than those of IoT. Accordingly, Intrusion Detection Systems (IDSs) have been developed to forestall inevitable harmful intrusions. IDSs survey the environment to identify intrusions in real time. This study designs an intrusion detection model exploiting feature engineering and machine learning for IIoT security. We combine Isolation Forest (IF) with Pearson's Correlation Coefficient (PCC) to reduce computational cost and prediction time. IF is exploited to detect and remove outliers from datasets. We apply PCC to choose the most appropriate features. PCC and IF are applied exchangeably (PCCIF and IFPCC). The Random Forest (RF) classifier is implemented to enhance IDS performances. For evaluation, we use the Bot-IoT and NF-UNSW-NB15-v2 datasets. RF-PCCIF and RF-IFPCC show noteworthy results with 99.98% and 99.99% Accuracy (ACC) and 6.18 s and 6.25 s prediction time on Bot-IoT, respectively. The two models also score 99.30% and 99.18% ACC and 6.71 s and 6.87 s prediction time on NF-UNSW-NB15-v2, respectively. Results prove that our designed model has several advantages and higher performance than related models.

Key words: Industrial Internet of Things (IIoT); isolation forest; Intrusion Detection Dystem (IDS); intrusion; Pearson's Correlation Coefficient (PCC); random forest

1 Introduction

The Internet of Things (IoT) represents an extensive scale network of integrated sensors and activators[1], serving a pertinent objective[2], and does not require human intervention arbitration[3–5]. Its emergence in numerous domains amplifies security questions[6]. Hence, further effort is required to deal with the newly created threats. Consequently, researchers suggest conventional tools to support resolving these issues[7–9]. In recent years, IoT technology has undergone quick evolution. Subsequently, its security is a mandatory task to warrant several services, such as confidentiality, privacy, data, and availability[10,11]. Due to the heterogeneous utilized protocols and data in IoT, implementing security mechanisms is becoming challenging[10]. Node quantity, low memory capability, processing command, and energy consumption have severely compromised security techniques[2,3]. The Industrial Internet of Things (IIoT) denotes linking devices, activators, and industrial systems to each. This technology gathers and analyses data to improve the industrial sector proficiency[12]. IIoT is measured as an IoT advancement

• Mouaad Mohy-Eddine, Azidine Guezzaz, and Said Benkirane are with the Technology Higher School, Cadi Ayyad University, Essaouira 44000, Morocco. E-mail: mouaadmohyeddine@gmail.com; a.guzzaz@gmail.com; sabenkirane@gmail.com.
• Mourade Azrour and Yousef Farhaoui are with the IDMS Team, Faculty of Sciences and Techniques, Moulay Ismail University of Meknès, Errachidia 52000, Morocco. E-mail: mo.azrour@umi.ac.ma; y.farhaoui@fste.umi.ac.ma.
* To whom correspondence should be addressed.
Manuscript received: 2022-06-25; revised: 2022-08-18; accepted: 2022-09-01

that aims to increase the computerization level via incorporating cloud and edge computing[13]. Similar to IoT, IIoT security has gained much attention lately[14]. IIoT security solutions aim to protect devices and transmitted data by proposing emerging programs and methods[15]. Industrial environments are mainly concerned with preventing replay, Denial of Service (DoS), Distributed DoS (DDoS), and Man-in-The-Middle (MiTM) attacks[15]. Therefore, Esfahani et al.[16] presented a mutual authentication to prevent replay and MiTM attacks. Yan et al.[17] proposed a multilevel DDoS mitigation framework to mitigate DDoS attacks in IIoT.

Moreover, an Intrusion Detection System (IDS) is implemented to monitor a host or system and detect normal intrusion instances[18]. IDSs rely on rules, signatures, states, or models to distinguish between normal and intrusion behaviors[19]. IDSs can be divided into signature, anomaly, and hybrid detection methods, which merge both to gain advantages[19]. IDSs are essential in maintaining networks from alterations and destructions[20,21]. Lately, Machine Learning (ML) has become a necessity for building well-performing IDSs. IDS methods have also captured zero-day attacks by adopting ML techniques[22,23]. Furthermore, ML methods have improved the Detection Rate (DR) and Accuracy (ACC) of IDS.

We propose and validate a Network IDS (NIDS) model for IIoT security. In our proposition, Isolation Forest (IF)[24] is integrated to achieve outlier detection. Pearson's Correlation Coefficient (PCC) is also applied to choose the most suitable features for dimensionality reduction. Then, Random Forest (RF)[25] distinguishes between normal packets and intrusions through binary classification. The obtained results indicate that our model is promising when it is compared with other previous related propositions. Our model depends on the Bot-IoT dataset, known for its imbalance, and on the NF-UNSW-NB15-v2 dataset. The strengths of our model appear in its capability to overtake the imbalance of the Bot-IoT dataset, especially when we remove the outliers and select the relevant features of the newly generated dataset.

The remainder of the paper is outlined below. Section 2 provides background on IoT and IIoT, IDSs, and ML. We also discuss related studies on IDSs, mainly works incorporating ML technologies. Section 3 introduces our suggested IDS model that is based on the RF classifier, dimensionality reduction, and feature selection. Section 4 details the implementation steps of our newly designed solution and shows the effectiveness of the proposed model through a discussion of the obtained results. Section 5 concludes and suggests future research topics.

2 Literature Survey

IoT adoption in the industrial sector has given birth to the new technology IIoT[15,26]. Hence, it can be defined as an extensive network connecting many sensors and actuators implemented in various fields, such as farming, healthcare, automotive, and smart grid[27]. IIoT presents different advantages to industrial fields to improve their efficiencies by allowing them to connect physical and virtual words. To connect to the Internet, IIoT can adopt various protocols, such as Message Queue Telemetry Transport and Long-Range Radio Wide Area Network[19]. IIoT architecture (Fig. 1) is similar to IoT architecture[28]. It comprises perception, networking, application, and cloud layers. The first layer refers to physical components[19,28]. The second layer contains communication protocols that transport data to the third layer, which merges, exploits, analyzes, and displays data to the end user[19,28]. The fourth layer provides scalability, storage, and comprehensive analytics[19,28].

Each IIoT layer is vulnerable to various threats and attacks. The most known ones are MiTM[16], DoS, DDoS[17], spoofing, and jamming attacks[14,15]. To deal with these attacks, various developers have implemented IDSs and other programs, such as access control, authentication systems[16], and encryption techniques[29]. Specifically, IDSs can be fixed and operated in any layer[19]. Typically, IDSs can be classified into three categories: Signature IDS (SIDS), Anomaly IDS (AIDS), and Hybrid IDS (HIDS). The first one (SIDS) can capture packets that are transmitted via the network. Once a packet is captured, it is compared to a database

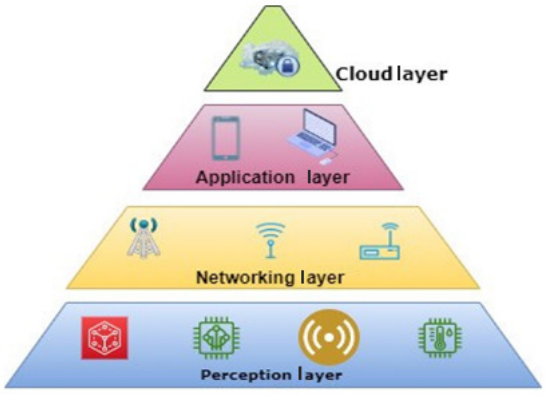


Fig. 1 IIoT architecture.

of recognized attacks[18,22,30,31]. Even so, only known attacks are detected with a low False Alarm Rate (FAR) and high DR. In addition, the requirement to frequently add new attacks to the database and the increasing zero-day attacks have made SIDS ineffective in various situations. The second one (AIDS) develops in response to the limits of SIDS. Hence, it can deduce the typical patterns and classify every irregular or nonconformity as an intrusion[30]. The third category (HIDS) is a combination of the first two[18]. Thus, it increases DR for known attacks and decreases FAR for unknown ones. Although other IDS types are discussed in the literature, for instance, host-based IDSs and Network IDSs (NIDSs)[18,22], we mentioned two types of IDS which we are going to define after. The host IDS is made for insider intrusion detection inside the host, and NIDS can capture and analyze all packets transmitted via the network[18,31]. NIDS cannot control high bandwidth and encrypted traffic[18]. Conversely, ML represents a subcategory of artificial intelligence, which attempts to provide machines and computers the ability to be further precise in predictions without requiring specific programming[32]. All ML algorithms have two key phases: training and testing. In the first phase, a model learns from the dataset the diverse patterns that lead to specific results. However, in the second phase, the model attempts to identify unidentified instances. ML algorithms can be divided into supervised and unsupervised learning[33]. Supervised classifiers are based on the pre-labeled data for training the model. Conventionally, different supervised classifiers are implemented to design IDSs, such as Decision Trees (DT)[34], k-nearest neighbors[35], Naïve Bayes (NB)[36], Support Vector Machine (SVM)[30,37], RF[38], and artificial neural network[39]. On the contrary, unsupervised learning uses unlabeled data in the training phase. Typically, various unsupervised models are used with IDSs, such as k-means[40–42], DBSCAN[43,44], and IF[45]. Nevertheless, RF[25] stands among the most popular supervised learning and central classifiers. The most significant characteristic of RF is its ability to process categorical values. RF can achieve high results in the classification case. It can combine multiple DTs with a bagging classifier, as it can use the RF[46] method to enhance obtained outcomes. In our classification case, the RF constructs multiple DTs in the training phase, and the RF result is the most selected feature by DTs.

Meanwhile, IF[24] is an unsupervised learning tree based anomaly detection algorithm. It is used to identify anomalous data occurrences by checking how a given point is distanced from the remainder of the dataset instead of investigating the regular items. IF operates effectively with large datasets, as it has a linear time complexity with a small memory overhead. The purpose of the procedure is to obtain an anomaly value for every subgroup of the dataset, measuring the divergence of the data in question. IF randomly chooses a point, and a descriptor then evaluates whether it can isolate the data. When the condition is reached, the algorithm stops. Otherwise, a new point and a new descriptor are picked randomly. The PCC algorithm determines the correlation between two series of data. PCC is used to calculate the correlation ratio between two variables. It outputs a value between -1 and 1, in which zero refers to the absence of correlation and $(1, -1)$ indicates a strong relationship.

Many related works have been proposed in the literature. Hence, Zhang et al.[47] proposed an IDS pretraining Wasserstein generative adversarial NIDS. They used LightGBM to double train the proposed model for detecting intrusions in IIoT networks and Wasserstein's generative adversarial network with gradient penalty. After testing the model, important results were obtained, that is, 99% F1-score and 100% accuracy when using the NSL-KDD dataset and 90% F1-score and 96% accuracy after using the CIC-IDS2018 dataset. Kasongo[19] designed IDS on the basis of RF, Linear Regression (LR), NB, DT, Extra Trees (ET), and extreme gradient boosting. They also used a Genetic Algorithm (GA) to select features and RF on the fitness function of the GA. For model validation, the UNSW-NB15 dataset was used. Hence, the model attained 87.61% accuracy and 0.98 on the Area Under the Curve (AUC). Furthermore, Raghuvanshi et al.[48] employed three algorithms (SVM, RF, and LR) to propose an IDS for smart farming. To validate their model, Raghuvanshi et al.[48] relied on the NSL-KDD dataset. The proposed models resulted in 98%, 85%, and 78% ACC to SVM, RF, and LR, respectively. Subsequently, Guezzaz et al.[49] designed an IDS model on the basis of DT and improved data quality on NSL-KDD and CIC-IDS2017 datasets. They compared the obtained results to related models using the same datasets. They proved that the suggested model had 99.42% and 98.8% ACC with NSL-KDD and CIC-IDS2017 datasets, respectively. Alhowaide et al.[50] proposed an IDS for IoT, which is based on an ensemble

learning model, and used the model selection method. They relied on NSL-KDD, UNSW-NB15, BoTNetIoT, and Bot-IoT datasets to evaluate the model. Therefore, the model achieved 99%, 95%, 100%, and 99% F1-scores and 100%, 98%, 100%, and 100% Receiver Operating Characteristic (ROC)-AUC scores when using NSL-KDD, UNSW-NB15, BoTNetIoT, and BoT-IoT datasets, respectively. Later, Javeed et al.[51] proposed Deep Learning (DL) Software-Defined Networking (SDN) -enabled smart framework dealing with IIoT area attacks. The Cu-LSTMGRU + Cu-BLSTM hybrid model was used to detect threats effectively. Hence, 99.45% ACC, 99.34% precision, 98.49% recall, and 99.47% F1-score were obtained. Afterward, Ge et al.[52] designed an IDS model for IoT by implementing DL. They developed a binary classifier (bFNN) and multiclass classifier (mFNN) to detect normal instances and other attack types. To evaluate their proposition, the BoT-IoT dataset was loaded. Hereafter, they obtained 99.99% ACC for mFNN and high ACC with few misclassified packets for bFNN. In Ref. [53], Malik et al. designed an NIDS for IoT traffic systems. They applied a Deep Belief Network (DBN) to perform the intrusion detection task.

Furthermore, they evaluated the DBN algorithm on a sample of the TON-IOT-Weather dataset and used a small number of epochs due to the limited computational power. Their model scored 86.33% ACC, 78% precision, 90% recall, and 84% F1-score. Alanazi and Aljuhani[54] designed an anomaly-based IDS that can reduce the risk of cyberattacks targeting IoT networks. The authors implemented the ensemble learning method in the detection phase and feature selection techniques for feature selection. The experimental outcomes on an online dataset were 99.984% ACC, 99.982% precision, 99.984% recall, and 99.983% F1-score. In Ref. [55], Lee et al. proposed a Multiclass classification based Intrusion Detection Model (M-IDM) in e-health IoT. The proposed model exploits data from healthcare sensors, such as electrocardiograms and thermometers. In this case, a conventional neural network classifies the traffic into multiple classes. M-IDM obtained 96.5%–96.7% AUC, 89%–93.7% F1-score, 91.1%–94.7% precision, and 84.4%–94.6% recall for 1×10^4, 5×10^4, and 1×10^5 instances. Maseer et al.[56] developed a hybrid DL IDS for IoT using a weighted DBN. The model integrates a Gaussian-Bernoulli restricted Boltzmann machine and a weighted deep neural network. They chose the CIC-IDS2017 dataset for evaluating the model. Hence, the obtained results were 99.38% and 99.99% ACC for web and bot attacks. Table 1 presents a summary of different models found in the literature.

3 Methodology

In this section, the implemented framework is described. Our proposition is based on the RF model and uses feature selection approaches to reduce time consumption. Hyperparameter tuning and performance measures are used to achieve the best performing settings for the proposed approach. Afterward, the newly created dataset

Table 1 A summary table of different models.

Reference	Year	Dataset	Model	ACC (%)
[47]	2021	NSL-KDD CIC-IDS2018	LightGBM and WGAN-GP	100 96
[19]	2021	UNSW-NB15	RF, LR, NB, DT, ET, and XGB	87.61
[48]	2022	NSL-KDD	SVM, RF, and linear regression	98, 85, and 78
[49]	2021	NSL-KDD CICIDS2017	DT	99.42 98.80
[50]	2021	NSL-KDD UNSW-NB15 BoTNetIoT BoT-IoT	MSM	–
[51]	2022	N-BaIoT	DL model	99.450
[52]	2021	BoT-IoT	bFNN mFNN	– 99.990
[53]	2022	TON-IOT-Weather	DBN	86.330
[54]	2022	Real-time dataset	Ensemble learning	99.984
[55]	2021	Real healthcare traffic	M-IDM	–
[56]	2021	CIC-IDS2017	Hybrid DL model	99.380

is utilized for training the RF classifier.

3.1 Proposed scheme

In general, IDS mechanisms comprise various modules, such as area data source, preprocessing, decision core, and response modules[29,30]. In this study, the implemented model is illustrated in Fig. 2. The main module of our proposed model is the preprocessing part. Hence, we take advantage of the PCC algorithm to select the most relevant features and use IF to detect outliers on the dataset.

We implement PCC as a feature reduction method to support model convergence, reduce computation cost and training time, and improve model performance[57–59], without any relevant data being lost. Furthermore, IF is used for detecting outliers on the Bot-IoT dataset to enhance the power of our model. Hereafter, the elimination of detected outliers significantly improves the model performance.

3.2 Solution description

As illustrated in Fig. 2, our suggested model comprises four phases: preprocessing, data quality, classifier training, and classification. We convert columns with string values to numerical ones in the preprocessing phase to improve the classifier speed. To deal with large values dominating the results, we use the Z-score normalization, which is the procedure of standardizing dataset values. Hence, the standard deviation is one, and the mean of all the data is 0. The formula used to calculate new values is detailed in Eq. (1).

$$\text{calculated value} = \frac{x - \mu}{\sigma} \tag{1}$$

where x represents the original value, μ refers to the data mean, and σ is the standard data deviation.

We adjust the model[60] to have the best settings using various hyperparameters. The execution of the model training phase on a prepared dataset shows the best performances. Formerly to decrease large numbers of datasets, we utilize PCC to look for less interrelated features and the strong correlation between features and the target value. PCC calculates the linear correlation between every two features and obtains a value between -1 and 1. According to Formula (2), we retain the features that produce in a PCC in the range $[0.5, -0.5]$ between each other or result in a PCC in the range $[1, 0.5]$ and $[-0.5, -1]$. The usage of PCC steps has a positive effect on the implementation of a well-organized classifier aimed at detecting intrusion.

$$\sigma_{\text{feature}} \Rightarrow \text{PCC}_{\text{feature}} \in \{([0.5, -0.5]_{(\text{feature, feature})}) \vee (([1, 0.5] \vee [-0.5, -1])_{(\text{feature, target})})\} \tag{2}$$

Our models are trained this way. RF is trained on the original datasets (called RF). We apply the IF outlier detector only on the datasets, then on RF (RF-IF).We performed RF after applying the PCC only (RF-PCC). RF trains on the generated data after IF, taking the PCC's output as input (RF-PCCIF). We perform PCC on the IF's output data (RF-IFPCC). We use 10-fold cross-validation as recommended in Ref. [37] to validate our proposed model. The 10-fold cross-validation randomly divides the dataset into ten fragments with identical sizes. Hence, 90% are served for model training, whereas 10% are used in the test phase. This process is iterated ten times to build an effective classifier that has the best performance and can detect new intrusions.

In the last part of our model, the purpose of the classifier is to give each instance the target value. Accordingly, the obtained model permits the detection of normal and abnormal instances. Thus, the RF algorithm is used to accomplish this job.

4 Experimental Study

For testing the proposed model's performance, we use the measures specified in Sections 4.2 and 4.3.

4.1 Dataset

In previously published works, we discover that several public datasets were used to evaluate ML-based IDSs[22]. In our case, Bot-IoT[59] and NF-UNSW-NB15-v2[61] datasets are loaded for training and validating the planned model. Details about the used datasets are given in the following subsections.

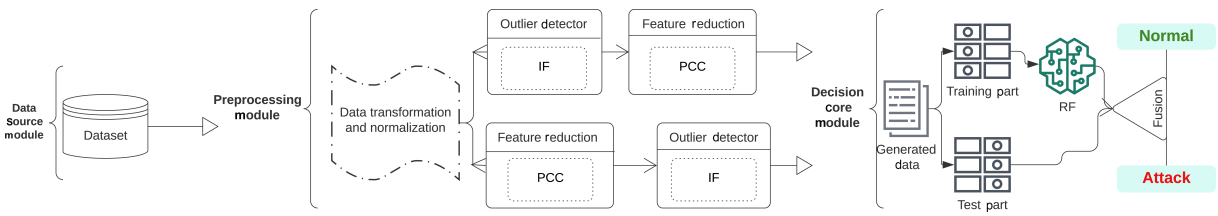


Fig. 2 Our proposed model architecture.

4.1.1 Bot-IoT dataset

Bot-IoT[59] is the most popular used dataset, mainly in research works related to IoT. This dataset is the fruit of Koroniotis et al.'s work. Koroniotis et al.[59] used a test bed to create Bot-IoT data in the Research Cyber Range Lab of UNSW Canberra. The database covers regular and attack examples with attack instance subgroups, such as DoS, DDoS, and service scanning. Hence, the overall saved packets are 73 370 443 instances subdivided into 9543 normal traffic and 73 360 900 attacks (Table 2). The generated dataset is available in various formats, such as CSV and Pcap files. In addition, certain files are disjointed according to classes. Table 2 presents the details about the used dataset, whereas Table 3 illustrates the 13 Bot-IoT features selected by PCC, the applied PCC on IF, and 15 by the applied output of IF on PCC.

4.1.2 NF-UNSW-NB15-v2 dataset

Initially, in 2015, the Research Cyber Range Lab of UNSW Canberra released the UNSW-NB15 dataset. Sarhan et al.[61] believed that the dataset suffers some limitations, such as dimensional overload and challenges in evaluating an ML model's generalization performance across several NIDS datasets using a specific or suggested feature set. Thus, they extracted a NetFlow version of the UNSW-NB15 called NF-UNSW-NB15. Two versions are available; the first version is made up of eight basic NetFlow features, whereas the second version is composed of 43 extended NetFlow features. In this study, we exploit the NF-UNSW-NB15-v2 dataset, which contains 1 623 118 instances in total, out of which 1 550 712 are normal instances and 72 406 are attacks. The dataset details are presented in Table 4.

Table 5 illustrates the NF-UNSW-NB15-v2 features selected by PCC, the applied PCC on IF, and the applied output of IF on PCC.

Table 2 Details about Bot-IoT.

Number of normal instances	Number of attacks	Total
9543	73 360 900	73 370 443

Table 3 List of the selected features from Bot-IoT.

Method	PCC	PCCIF	IFPCC
Selected feature	stime	stime	stime
	flgs_number	flgs_number	flgs_number
	saddr	saddr	saddr
	daddr	daddr	sport
	pkts	pkts	pkts
	state_number	state_number	dport
	seq	seq	bytes
	dur	dur	sum
	stddev	stddev	min
	min	min	spkts
	rate	rate	dpkts
	srate	srate	sbytes
	drate	drate	TnBPSrcIP
	–	–	TnP_PDstIP
	–	–	TnP_Per_Dport

4.2 Experiment evaluation

Experimental studies are executed on our personal computer, which has the following characteristics: an Intel (R) Core (TM) i5-6200U CPU@2.30 GHz and 12 GB DDR3 on RAM. The installed operating system is Windows 10 Pro x64-bit. The implementation of our proposed model and feature engineering is accomplished under Python v3.9.6.

4.3 Performance metrics

The performance of the proposed model is described with the usage of usual metrics, such as accuracy, precision, recall, F1-score, AUC, and False Positive Rate (FPR).

- Accuracy refers to the ratio of the truly classified data over all instances of the dataset. It is computed according to Eq. (3).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \tag{3}$$

- Precision is the data that are truly detected as attacks divided by the sum of normal instances and attacks detected as attacks.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \tag{4}$$

- Recall (True Positive Rate (TPR)) represents True Positives (TPs) divided by the total of TPs and false positives.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \tag{5}$$

- F1-score denotes the harmonic mean of the two previous metrics: recall and precision.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \tag{6}$$

- False Positive Rate (FPR) mentions the ratio of wrongly classified attacks over actual normal instances.

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \times 100\% \tag{7}$$

Table 4 Statistics of the NF-UNSW-NB15-v2.

Number of normal instances	Number of attacks	Total
1 550 712	72 406	1 623 118

Table 5 List of selected features from NF-UNSW-NB15-v2.

Method	PCC	PCCIF	IFPCC
Selected feature	ipv4_src_addr	ipv4_src_addr	ipv4_src_addr
	l4_src_port	l4_src_port	l4_src_port
	ipv4_dst_addr	ipv4_dst_addr	ipv4_dst_addr
	l4_dst_port	l4_dst_port	l4_dst_port
	protocol	protocol	17_proto
	17_proto	17_proto	in_bytes
	in_bytes	in_bytes	flow_duration_milliseconds
	out_bytes	out_bytes	min_ip_pkt_len
	duration_out	duration_out	max_ip_pkt_len
	min_ttl	min_ttl	src_to_dst_second_bytes
	longest_flow_pkt	longest_flow_pkt	dst_to_src_second_bytes
	shortest_flow_pkt	shortest_flow_pkt	src_to_dst_avg_throughput
	src_to_dst_second_bytes	src_to_dst_second_bytes	dst_to_src_avg_throughput
	dst_to_src_second_bytes	dst_to_src_second_bytes	num_pkts_up_to_128_bytes
	src_to_dst_avg_throughput	src_to_dst_avg_throughput	num_pkts_128_to_256_bytes
	dst_to_src_avg_throughput	dst_to_src_avg_throughput	num_pkts_256_to_512_bytes
	num_pkts_128_to_256_bytes	num_pkts_128_to_256_bytes	num_pkts_512_to_1024_bytes
	num_pkts_256_to_512_bytes	num_pkts_256_to_512_bytes	num_pkts_1024_to_1514_bytes
	tcp_win_max_out	tcp_win_max_out	tcp_win_max_out
	icmp_type	icmp_type	icmp_type
	dns_query_type	dns_query_type	dns_query_id
	dns_ttl_answer	dns_ttl_answer	dns_query_type
	ftp_command_ret_code	ftp_command_ret_code	dns_ttl_answer
	dns_query_id	dns_query_id	ftp_command_ret_code

• AUC is the classifier ability point to distinguish among classes.

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}(\mathrm{FPR}(t))\mathrm{d}\,\mathrm{FPR}(t) \tag{8}$$

4.4 Result discussion

4.4.1 Bot-IoT dataset result discussion

Table 6 shows the used metrics to evaluate our proposed model. Figure 3 compares the model outcomes on the basis of these metrics. As displayed in Fig. 3, we determine that the five classifiers on the Bot-IoT dataset have similar ACC and precision results, with 99.99% for RF, RF-IF, and RF-IFPCC and 99.98% for RF-PCC and RF-PCCIF. A 100% recall and 99.99% F1-score are observed for all. The classifiers display considerable differences in ROC scores with 71.28% RF, 89.86% RF-IF, 59.9% RF-PCC, 60.78% RF-PCCIF, and as the best performer, RF-IFPCC scores 92.48%.

As detailed above, the usage of IF and PCC can reduce time costs considerably without affecting the model performance. Accordingly, and as shown in Fig. 4, the projected models score 717.51 s, 344.20 s, 484.40 s, 217.64 s, and 210.73 s on RF, RF-IF, RF-PCC, RF-PCCIF, and RF-IFPCC, respectively. They score 8.93 s, 7.17 s, 10.14 s, 6.18 s, and 6.25 s on prediction time on RF, RF-IF, RF-PCC, RF-PCCIF, and RF-IFPCC,

Table 6 A summary table of performances metrics on Bot-IoT and NF-UNSW-NB15-v2 datasets.

Dataset	Model	ACC (%)	Precision (%)	Recall (%)	F1-score (%)	ROC score (%)	Training time (s)	Prediction time (s)
Bot-IoT	RF	99.99	99.99	100.00	99.99	71.28	717.51	8.93
	RF-IF	99.99	99.99	100.00	99.99	89.86	344.20	7.17
	RF-PCC	99.98	99.98	100.00	99.99	59.90	484.40	10.14
	RF-PCCIF	99.98	99.98	100.00	99.99	60.78	217.64	6.18
	RF-IFPCC	99.99	99.99	100.00	99.99	92.48	210.73	6.25
NF-UNSW-NB15-v2	RF	99.17	82.96	99.75	90.58	99.45	251.57	5.97
	RF-IF	99.18	82.97	99.78	90.60	99.47	189.50	7.58
	RF-PCC	99.27	84.60	99.91	91.62	99.58	173.53	5.20
	RF-PCCIF	99.30	85.18	99.87	91.94	99.58	145.24	6.71
	RF-IFPCC	99.18	83.20	99.61	90.67	99.39	146.44	6.87

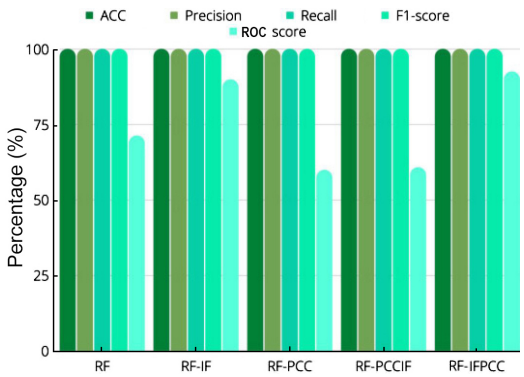


Fig. 3 Summary of performance metrics on the Bot-IoT dataset.

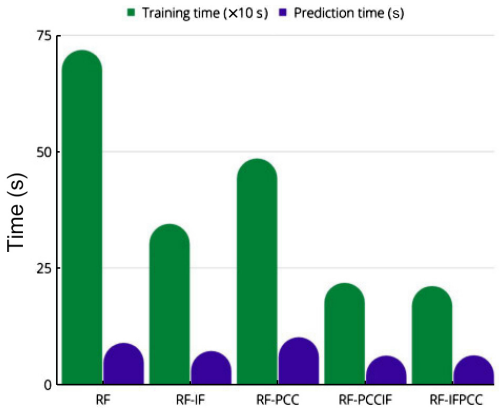


Fig. 4 Training time and prediction time consumed by our models on Bot-IoT.

respectively.

Figure 4 illustrates the training and prediction time consumed by our models on Bot-IoT.

The RF confusion matrix and its ROC curve are illustrated in Fig. 5. The RF model can predict intrusions with 100% TP. However, the RF classifier does not function well in predicting True Negatives (TNs) (43%) and False Negatives (FNs) (57%). Furthermore, the ROC curve demonstrates the aptitude of a classifier to differentiate between normal and attack instances. The curve is obtained by plotting the TPR against the FPR. The RF ROC curve displays the high distinguishing capability of the RF model.

Figure 6 displays the confusion matrix of the RF-IF model and its ROC curve. As displayed, the RF-IF classifier can predict significant results, with 80% TN and only 20% FP. In addition, it retains the identical

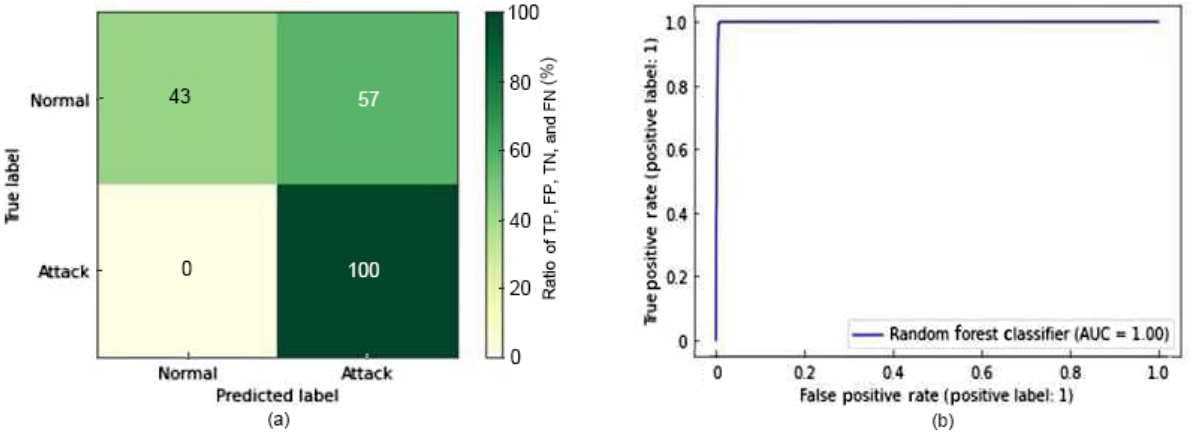


Fig. 5 Confusion matrix and ROC curve of RF on the Bot-IoT dataset.

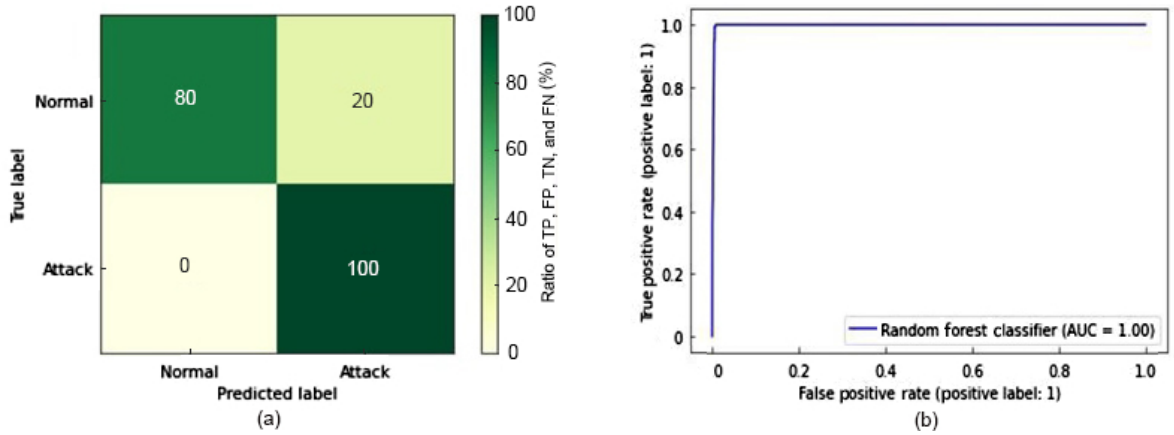


Fig. 6 Confusion matrix and ROC curve of RF-IF on the Bot-IoT dataset.

result as the RF model when we see TP with 100%. Hence, the RF-IF's ROC curve shows the model's ability to distinguish among classes.

Similarly, Fig. 7 shows the confusion matrix of the RF-PCC model and its ROC curve. This confusion matrix demonstrates that the model misclassifies normal instances. Nevertheless, it works in the case of the prediction of attack instances with 100% TP. In addition, the ROC curve displays high performance in forecasting attacks.

Figure 8 illustrates the confusion matrix of the RF-PCCIF model and its ROC curve. One can observe that the model does not efficiently predict normal instances with 78% in FN. Nevertheless, it shows significant importance in the prediction of attack instances. The ROC curve confirms the considerable performance of predicting attacks.

Figure 9 shows the confusion matrix of the RF-IFPCC

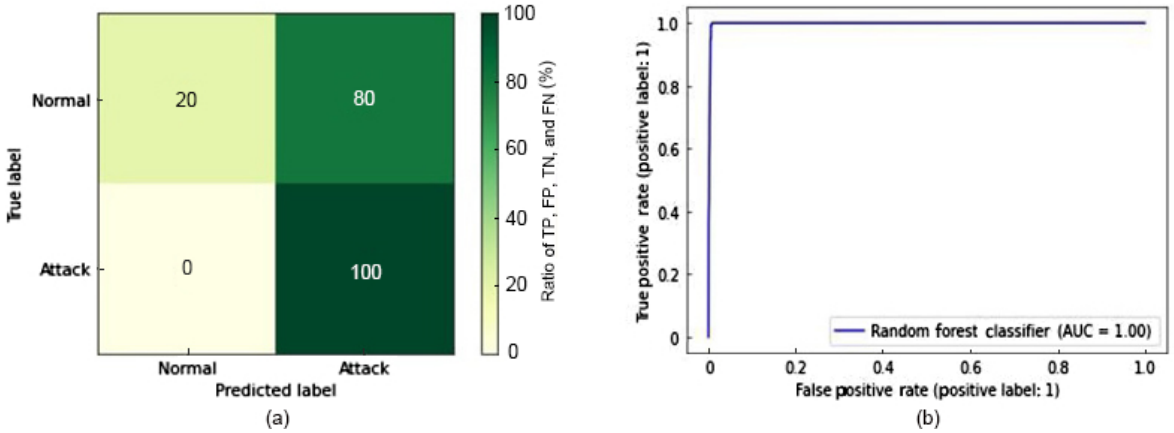


Fig. 7 Confusion matrix and ROC curve of RF-PCC on the Bot-IoT dataset.

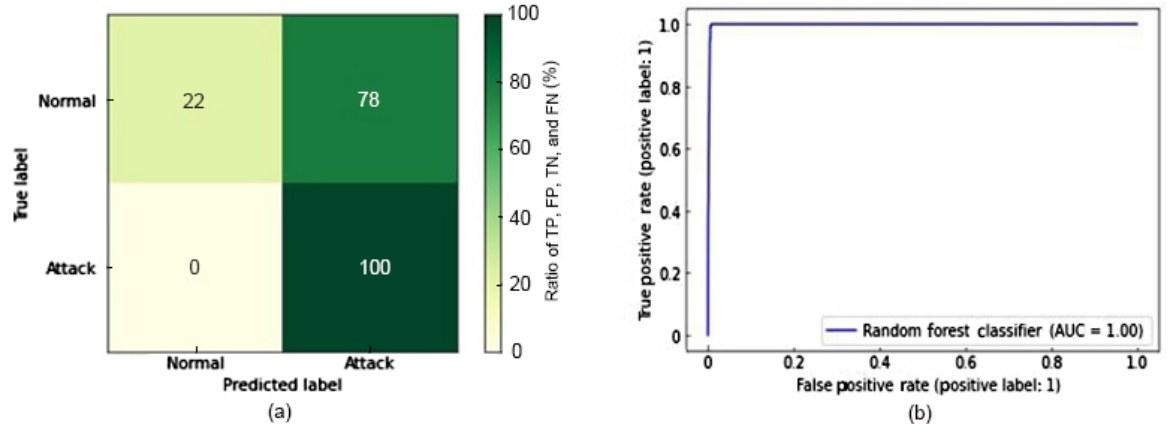


Fig. 8 Confusion matrix and ROC curve of RF-PCCIF on the Bot-IoT dataset.

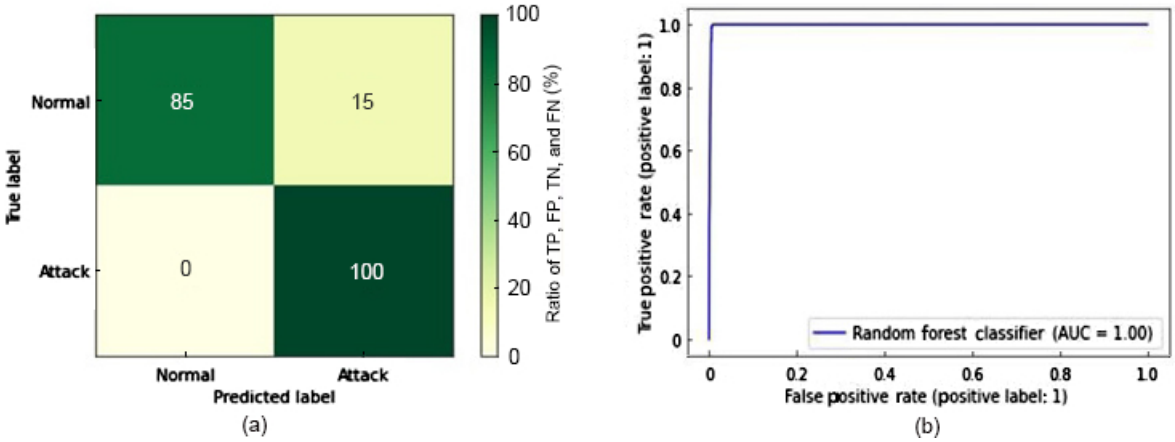


Fig. 9 Confusion matrix and ROC curve of RF-IFPCC on the Bot-IoT dataset.

model and its ROC curve. We observe that the model performs well in predicting normal instances, with 85% in TN. Moreover, it presents great predictions on attack instances. The ROC curve confirms the significant performances of this model in distinguishing normal and attack instances, unlike previous models.

4.4.2 NF-UNSW-NB15-v2 dataset result discussion

Table 6 presents the used metrics to evaluate our proposed model. Figure 10 illustrates the metrics applied to evaluate our models and compare the results. By observing Table 6 and Fig. 10, the five classifiers on the NF-UNSW-NB15-v2 dataset have significant results on ACC with 99.17%, 99.18%, 99.27%, 99.30%, and 99.18% for RF, RFIF, RF-PCC, RF-PCCIF, and RF-IFPCC, respectively. On precision, the models score 82.96%, 82.97%, 84.6%, 85.18%, and 83.2% for RF, RF-IF, RF-PCC, RF-PCCIF, and RF-IFPCC, respectively. On recall, 99.75%, 99.78%, 99.91%, 99.87%, and 99.61% are provided for RF, RF-IF, RF-PCC, RF-PCCIF, and RF-IFPCC, respectively.

On F1-score, 90.58%, 90.6%, 91.62%, 91.94%, and 90.67% are given for RF, RF-IF, RF-PCC, RF-PCCIF, and RF-IFPCC, respectively. The classifiers almost have the same ROC score with 99.45% RF, 99.47% RF-IF, 99.58% RF-PCC, 99.58% RF-PCCIF, and 99.39% RF-IFPCC.

As displayed in Table 6 and Fig. 11, the models score 235.17 s, 189.50 s, 173.53 s, 145.24 s, and 146.44 s on RF, RF-IF, RF-PCC, RF-PCCIF, and RF-IFPCC, respectively. They score 5.67 s, 7.58 s, 5.20 s, 6.71 s, and 6.87 s on prediction time for RF, RF-IF, RF-PCC, RF-PCCIF, and RF-IFPCC, respectively.

Figure 12 illustrates the confusion matrix of the RF model and its ROC curve on the NF-UNSW-NB15-v2 dataset. It shows that the intrusion prediction rate of the RF model is 99.75% TP. On TN, the RF classifier scores 99.15%, with only 0.85% FN. ROC curve presents the

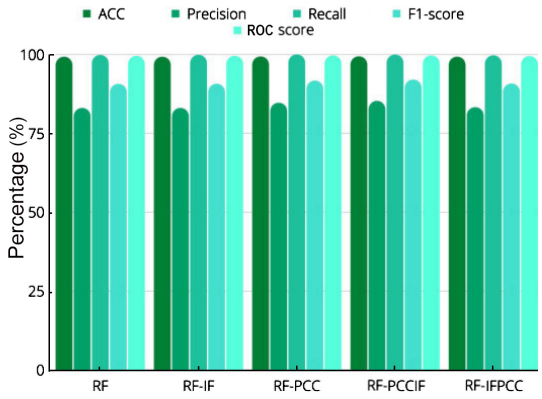


Fig. 10 Summary of performance metrics on the NF-UNSW-NB15-v2 dataset.

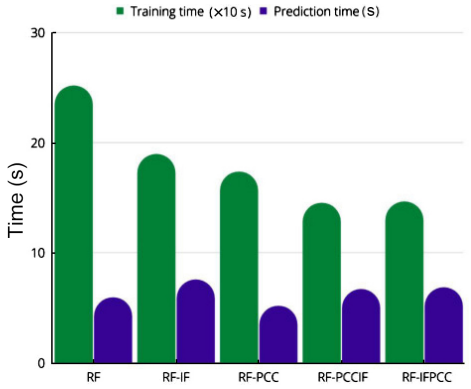


Fig. 11 Training time and prediction time consumed by our models on the NF-UNSW-NB15-v2 dataset.

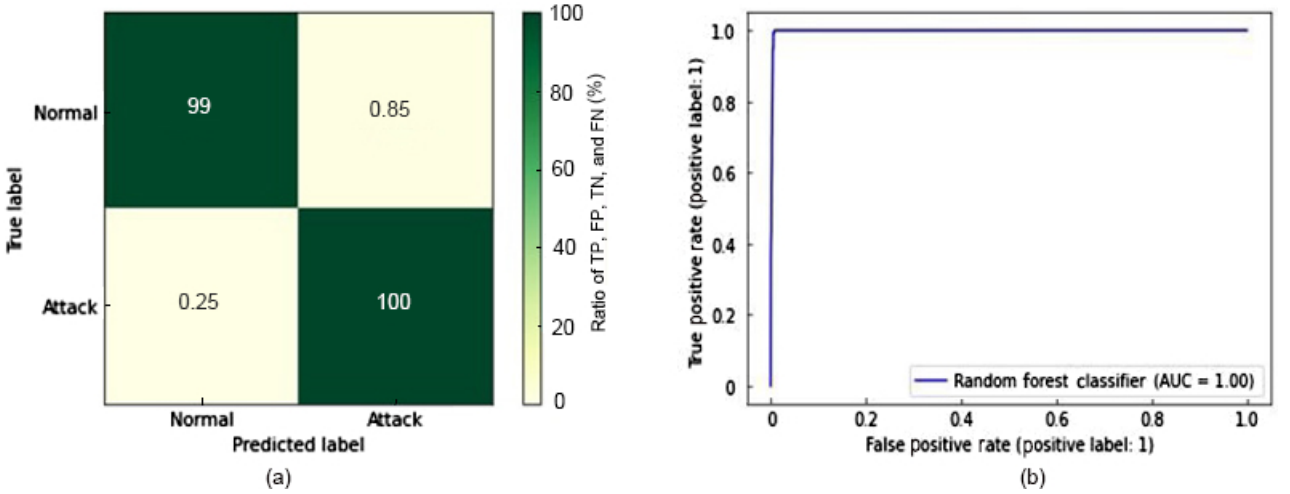


Fig. 12 Confusion matrix and ROC curve of RF on the NF-UNSW-NB15-v2 dataset.

capacity of a classifier to distinguish between normal and attack instances. We can obtain the curve by plotting the TPR against the FPR. The RF ROC curve shows the high distinguish capacity of the RF model.

Figure 13 presents the confusion matrix of the RF-IF model and its ROC curve on the NF-UNSW-NB15-v2 dataset. RF-IF exhibits performance close to the RF model with 99.16% TN and 0.84% FN. Moreover, it shows a similar result as the RF model regarding TP with 99.79%. Similar to the RF ROC curve, the RF-IF ROC curve shows the same distinguishing capability.

Figure 14 presents the confusion matrix of the RF-PCC model and its ROC curve on the NF-UNSW-NB15-v2 dataset. The confusion matrix shows that the model misclassifies 0.75% of the normal instances. Nevertheless, it performs well in predicting attack instances with 99.92% TP. Moreover, the ROC curve accurately predicts attacks and normal instances.

Figure 15 depicts the confusion matrix of the RF-PCCIF model and its ROC curve on the NF-UNSW-NB15-v2 dataset. The confusion matrix presents that the model correctly classifies 99.28% of the normal instances. Furthermore, it performs well in predicting attack instances with 99.88% TP. The ROC curve shows that this model performs well in distinguishing normal from attack instances.

Figure 16 illustrates the confusion matrix of the RF-IFPCC model and its ROC curve on the NF-UNSW-NB15-v2 dataset. Examining the confusion matrix, we deduce that the model can classify 99.17% of the normal instances well. The model performs well in predicting attack instances with 99.62%. The ROC curve demonstrates that the model has an excellent performance in predicting instances correctly.

On the overall performance, the RF-IFPCC classifier generates the most effective results on the Bot-IoT

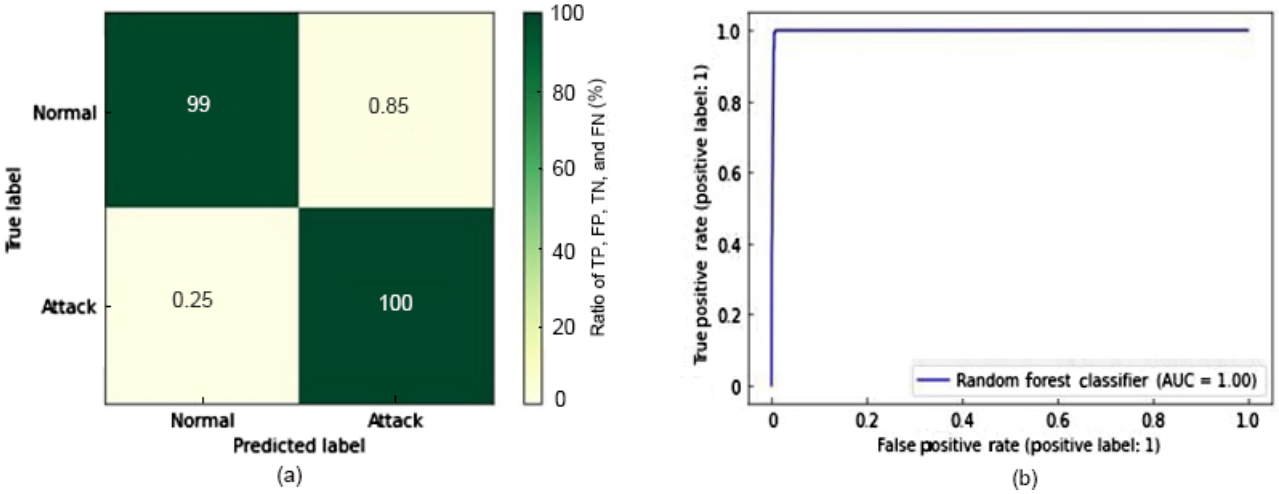


Fig. 13 Confusion matrix and ROC curve of RF-IF on the NF-UNSW-NB15-v2 dataset.

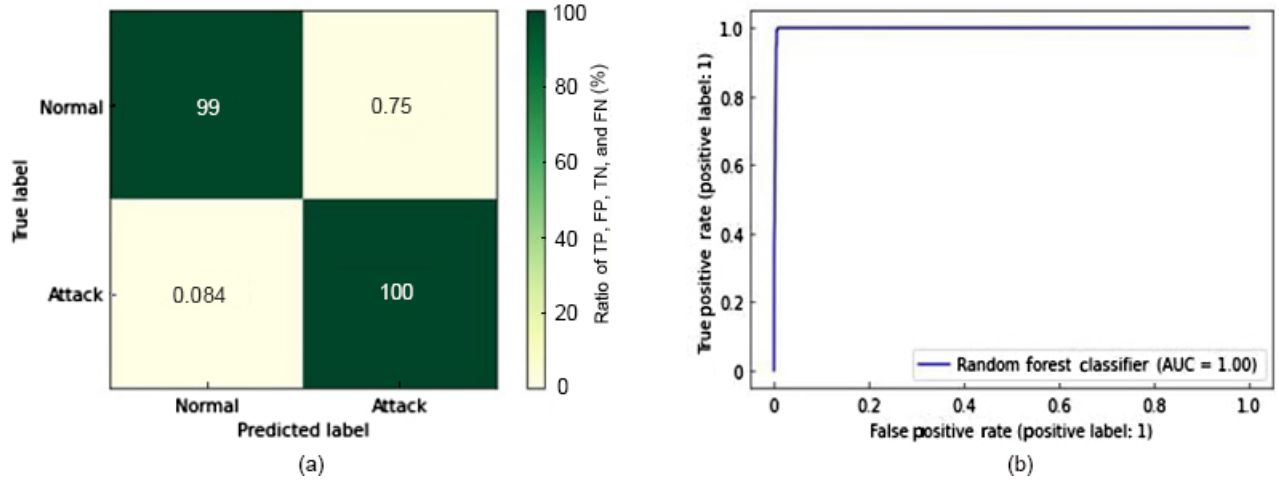


Fig. 14 Confusion matrix and ROC curve of RF-PCC on the NF-UNSW-NB15-v2 dataset.

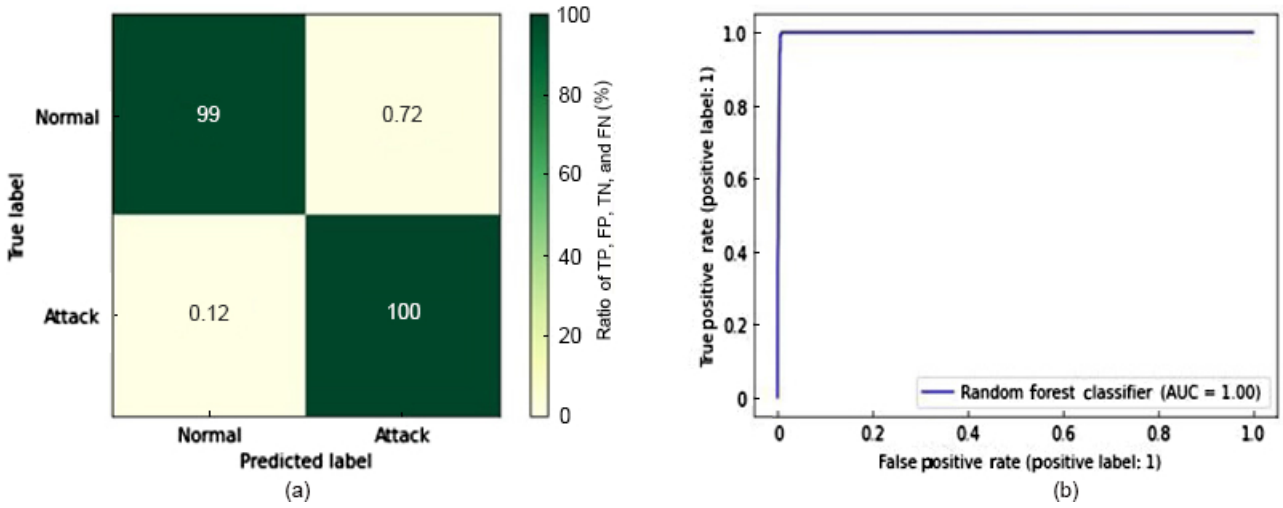


Fig. 15 Confusion matrix and ROC curve of RF-PCCIF on the NF-UNSW-NB15-v2 dataset.

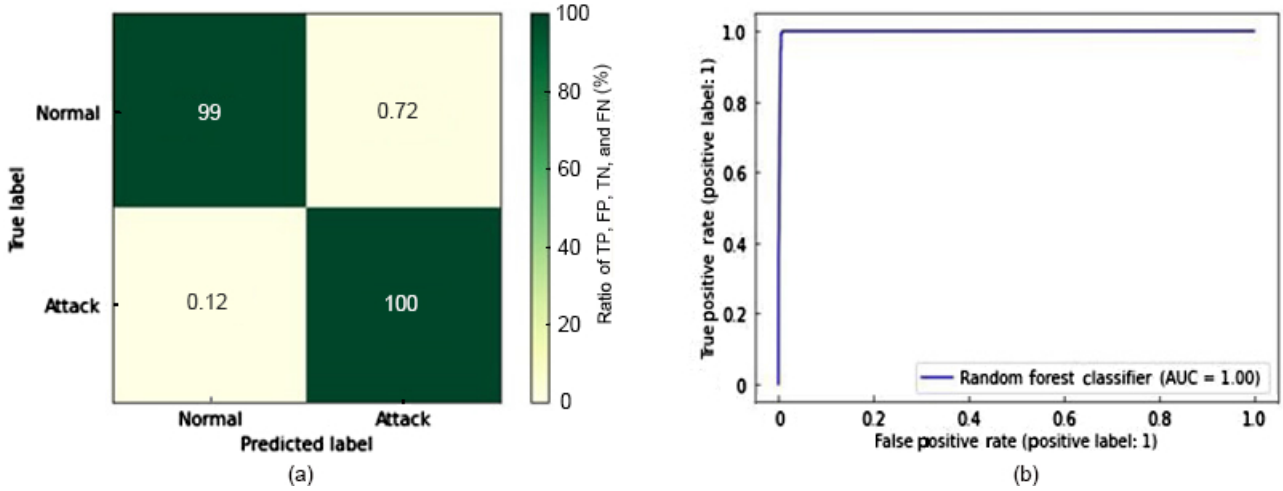


Fig. 16 Confusion matrix and ROC curve of RF-IFPCC on the NF-UNSW-NB15-v2 dataset.

dataset. Our model goes from detecting only 43% of normal attacks to correctly detecting over 85% of them, proving that our model overtakes the infamous imbalance of the Bot-IoT dataset without influencing its ability to detect severe intrusions. Meanwhile, RF-IFPCC and RF-PCCIF scores are approximately the same, outperforming the other methods on the NF-UNSW-NB15-v2 dataset. As presented in Table 6, the models exhibit better performances than those in previous research that used the same datasets.

Our models are evaluated on Bot-IoT and NF-UNSW-NB15-v2 datasets and are improved through feature selection and dimensionality reduction methods. Thus, convenient results are shown compared with other approaches in the literature, as presented in Table 7.

5 Conclusion

The proven efficiency of IDS has made it an essential

Table 7 Performance comparison between our model and previous works on Bot-IoT and NF-USNW-NB15 datasets.

Dataset	Paper	Year	Model	Feature selection	ACC (%)	F1-score (%)
Bot-IoT	Nimbalkar and Kshirsagar[62]	2021	Information gain and gain ratio	–	99.99	–
	Abushwereb et al.[63]	2022	Chi-square	–	50.90, 99.50, 99.70	–
	Saba et al.[64]	2022	CNN	–	92.85	–
	Our approach	2022	RF	IF + PCC	99.99	99.99
NF-UNSW-NB15-v2	Sarhan et al.[61]	2022	ET	–	99.73	97.00
	Our approach	2022	RF	PCC + IF	99.30	91.94

tool, among others, to mitigate IIoT vulnerabilities. In this study, we implement an IDS for IIoT networks using the RF model for classification, PCC to select relevant features, and IF as an outlier detector. We use PCC and IF separately and exchangeably. IF takes the output of PCC as input and vice versa. Our model shows a great result in overtaking the imbalance of the Bot-IoT dataset, which can be seen in the confusion matrix of the RF-IFPCC model. On the NF-UNSW-NB15-v2 dataset, the results are close to one another with outstanding performances. We intend in our future work to exploit other datasets, such as the TON-IoT dataset containing IoT and IIoT data, to have a global view and create and validate an effective IDS for improving network security in general.

References

[1] P. M. Chanal and M. S. Kakkasageri, Security and privacy in IoT: A survey, *Wireless Personal Communications*, vol. 115, pp. 1667–1693, 2020.

[2] P. Sethi and S. R. Sarangi, Internet of things: Architectures, protocols, and applications, *Journal of Electrical and Computer Engineering*, vol. 2017, p. 9324035, 2017.

[3] F. A. Alaba, M. Othman, I. A. T. Hashem, and F. Alotaibi, Internet of things security: A survey, *Journal of Network and Computer Applications*, vol. 88, pp. 10–28, 2017.

[4] M. Azrour, J. Mabrouki, Y. Farhaoui, and A. Guezzaz, Security analysis of Nikooghadam et al.'s authentication protocol for cloud-IoT, in *Intelligent Systems in Big Data, Semantic Web and Machine Learning*, N. Gherabi and J. Kacprzyk, eds. Cham, Switzerland: Springer, 2021, pp. 261–269.

[5] M. Moutaib, T. Ahajjam, M. Fattah, Y. Farhaoui, B. Aghoutane, and M. E. Bekkali, Application of internet of things in the health sector: Toward minimizing energy consumption, *Big Data Mining and Analytics*, vol. 5, no. 4, pp. 302–308, 2022.

[6] M. Azrour, J. Mabrouki, A. Guezzaz, and Y. Farhaoui, New enhanced authentication protocol for internet of things, *Big Data Mining and Analytics*, vol. 4, no. 1, pp. 1–9, 2021.

[7] R. V. Solms and J. V. Niekerk, From information security to cyber security, *Computers & Security*, vol. 38, pp. 97–102, 2013.

[8] M. Azrour, J. Mabrouki, A. Guezzaz, and A. Kanwal, Internet of things security: Challenges and key issues, *Security and Communication Networks*, vol. 2021, p. 5533843, 2021.

[9] A. Guezzaz, S. Benkirane, and M. Azrour, A novel anomaly network intrusion detection system for internet of things security, in *IoT and Smart Devices for Sustainable Environment*, M. Azrour, A. Irshad, and R. Chaganti, eds. Cham, Switzerland: Springer, 2022, pp. 129–138.

[10] M. B. M. Noor and W. H. Hassan, Current research on internet of things (IoT) security: A survey, *Computer Networks*, vol. 148, pp. 283–294, 2019.

[11] M. A. Khan, M. A. K. Khattk, S. Latif, A. A. Shah, M. U. Rehman, W. Boulila, M. Driss, and J. Ahmad, Voting classifier-based intrusion detection for IoT networks, in *Advances on Smart and Soft Computing*, F. Saeed, T. Al-Hadhrami, E. Mohammed, and M. Al-Sarem, eds. Singapore: Springer, 2022, pp. 313–328.

[12] X. Yu and H. Guo, A survey on IIoT security, in *Proc. 2019 IEEE VTS Asia Pacific Wireless Communications Symposium* (*APWCS*), Singapore, 2019, pp. 1–5.

[13] K. Tange, M. D. Donno, X. Fafoutis, and N. Dragoni, A systematic survey of industrial internet of things security: Requirements and fog computing opportunities, *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2489–2520, 2020.

[14] A. C. Panchal, V. M. Khadse, and P. N. Mahalle, Security issues in IIoT: A comprehensive survey of attacks on IIoT and its countermeasures, in *Proc. 2018 IEEE Global Conference on Wireless Computing and Networkin*g (*GCWCN*), Lonavala, India, 2018, pp. 124–130.

[15] J. Sengupta, S. Ruj, and S. D. Bit, A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoT, *Journal of Network and Computer Applications*, vol. 149, p. 102481, 2020.

[16] A. Esfahani, G. Mantas, R. Matischek, F. B. Saghezchi, J. Rodriguez, A. Bicaku, S. Maksuti, M. G. Tauber, C. Schmittner, and J. Bastos, A lightweight authentication mechanism for M2M communications in industrial IoT environment, *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 288–296, 2019.

[17] Q. Yan, W. Huang, X. Luo, Q. Gong, and F. R. Yu, A multi-level DDoS mitigation framework for the industrial internet of things, *IEEE Communications Magazine*, vol. 56, no. 2, pp. 30–36, 2018.

[18] A. L. Buczak and E. Guven, A survey of data mining and machine learning methods for cyber security intrusion detection, *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016.

[19] S. M. Kasongo, An advanced intrusion detection system for IIoT based on GA and tree based algorithms, *IEEE Access*, vol. 9, pp. 113199–113212, 2021.

[20] A. Aldweesh, A. Derhab, and A. Z. Emam, Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues, *Knowledge-Based Systems*, vol. 189, p. 105124, 2020.

[21] A. Guezzaz, A. Asimi, Y. Asimi, Z. Tbatou, and Y. Sadqi, A global intrusion detection system using PcapSockS sniffer and multilayer perceptron classifier, *Int. J. Netw. Secur.*, vol. 21, no. 3, pp. 438–450, 2019.

[22] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, Survey of intrusion detection system: Techniques, datasets and challenges, *Cybersecurity*, vol. 2, p. 20, 2019.

[23] A. Guezzaz, Y. Asimi, M. Azrour, and A. Asimi, Mathematical validation of proposed machine learning classifier for heterogeneous traffic and anomaly detection, *Big Data Mining and Analytics*, vol. 4, no. 1, pp. 18–24, 2021.

[24] F. T. Liu, K. M. Ting, and Z. -H. Zhou, Isolation forest, in *Proc. 2008 Eighth IEEE International Conference on Data Mining*, Pisa, Italy, 2008, pp. 413–422.

[25] T. K. Ho, Random decision forests, in *Proc. 3rd International Conference on Document Analysis and Recognition*, Montreal, Canada, 1995, pp. 278–282.

[26] T. Ainsworth, J. Brake, P. Gonzalez, D. Toma, and A. F. Browne, A comprehensive survey of industry 4.0, IIOT and areas of implementation, in *Proc. SoutheastCon 2021*, Atlanta, GA, USA, 2021, pp. 1–6.

[27] P. K. Malik, R. Sharma, R. Singh, A. Gehlot, S. C. Satapathy, W. S. Alnumay, D. Pelusi, U. Ghosh, and J. Nayak, Industrial internet of things and its applications in industry 4.0: State of the art, *Computer Communications*, vol. 166, pp. 125–139, 2021.

[28] L. Hylving and U. Schultze, Evolving the modular layered architecture in digital innovation: The case of the car's instrument cluster, presented at 34th International Conference on Information Systems, Milan, Italy, 2013.

[29] M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study, *Journal of Information Security and Applications*, vol. 50, p. 102419, 2020.

[30] J. Gu and S. Lu, An effective intrusion detection approach using SVM with Naïve Bayes feature embedding, *Computers & Security*, vol. 103, p. 102158, 2020.

[31] H. -J. Liao, C. -H. R. Lin, Y. -C. Lin, and K. -Y. Tung, Intrusion detection system: A comprehensive review, *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 16–24, 2013.

[32] M. Azrour, J. Mabrouki, G. Fattah, A. Guezzaz, and F. Aziz, Machine learning algorithms for efficient water quality prediction, *Modeling Earth Systems and Environment*, vol. 8, no. 2, pp. 2793–2801, 2022.

[33] A. K. Sandhu, Big data with cloud computing: Discussions and challenges, *Big Data Mining and Analytics*, vol. 5, no. 1, pp. 32–40, 2021.

[34] K. Peng, V. C. M. Leung, L. Zheng, S. Wang, C. Huang, and T. Lin, Intrusion detection system based on decision tree over big data in fog environment, *Wireless Communications and Mobile Computing*, vol. 2018, p. 4680867, 2018.

[35] R. Wazirali, An improved intrusion detection system based on KNN hyperparameter tuning and cross-validation, *Arabian Journal for Science and Engineering*, vol. 45, no. 12, pp. 10859–10873, 2020.

[36] A. A. Sallam, M. N. Kabir, Y. M. Alginahi, A. Jamal, and T. K. Esmeel, IDS for improving DDoS attack recognition based on attack profiles and network traffic features, in *Proc. 2020 16th IEEE International Colloquium on Signal Processing & Its Applications* (*CSPA*), Langkawi, Malaysia, 2020, pp. 255–260.

[37] J. Gu, L. Wang, H. Wang, and S. Wang, A novel approach to intrusion detection using SVM ensemble with feature augmentation, *Computers & Security*, vol. 86, pp. 53–62, 2019.

[38] S. Waskle, L. Parashar, and U. Singh, Intrusion detection system using PCA with random forest approach, in *Proc. 2020 International Conference on Electronics and Sustainable Communication Systems* (*ICESC*), Coimbatore, India, 2020, pp. 803–808.

[39] J. O. Mebawondu, O. D. Alowolodu, J. O. Mebawondu, and A. O. Adetunmbi, Network intrusion detection system using supervised learning paradigm, *Scientific African*, vol. 9, p. e00497, 2020.

[40] J. Doe, The dictionary of substances and their effects, http://www.rsc.org/dose/title, 1999.

[41] J. Chen, X. Qi, L. Chen, F. Chen, and G. Cheng, Quantum-inspired ant lion optimized hybrid k-means for cluster analysis and intrusion detection, *Knowledge-Based Systems*, vol. 203, p. 106167, 2020.

[42] L. Wang, J. Yang, M. Workman, and P. Wan, Effective algorithms to detect stepping-stone intrusion by removing outliers of packet RTTs, *Tsinghua Science and Technology*, vol. 27, no. 2, pp. 432–442, 2021.

[43] A. Saxena, K. Saxena, and J. Goyal, Hybrid technique based on DBSCAN for selection of improved features for intrusion detection system, in *Emerging Trends in Expert Applications and Security*, V. S. Rathore, M. Worring, D. K. Mishra, A. Joshi, S. Maheshwari, eds. Singapore: Springer, 2019, pp. 365–377.

[44] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in *Proc. Second International Conference on Knowledge Discovery and Data Mining*, Portland, OR, USA, 1996, pp. 226–231.

[45] K. Sadaf and J. Sultana, Intrusion detection based on autoencoder and isolation forest in fog computing, *IEEE Access*, vol. 8, pp. 167059–167068, 2020.

[46] A. Sarica, A. Cerasa, and A. Quattrone, Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review, *Frontiers in Aging Neuroscience*, vol. 9, p. 329, 2017.

[47] L. Zhang, S. Jiang, X. Shen, B. B. Gupta, and Z. Tian, PWG-IDS: An intrusion detection model for solving class imbalance in IIoT networks using generative adversarial networks, arXiv preprint arXiv: 2110.03445, 2021.

[48] A. Raghuvanshi, U. K. Singh, G. S. Sajja, H. Pallathadka, E. Asenso, M. Kamal, A. Singh, and K. Phasinam, Intrusion detection using machine learning for risk mitigation in IoT-enabled smart irrigation in smart farming, *Journal of Food Quality*, vol. 2022, p. 3955514, 2022.

[49] A. Guezzaz, S. Benkirane, M. Azrour, and S. Khurram, A reliable network intrusion detection approach using decision tree with enhanced data quality, *Security and Communication Networks*, vol. 2021, p. 1230593, 2021.

[50] A. Alhowaide, I. Alsmadi, and J. Tang, Ensemble detection model for IoT IDS, *Internet of Things*, vol. 16, p. 100435, 2021.

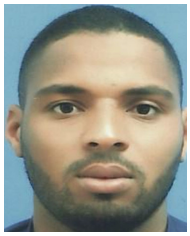
[51] D. Javeed, T. Gao, M. T. Khan, and D. Shoukat, A hybrid intelligent framework to combat sophisticated threats in secure industries, *Sensors*, vol. 22, no. 4, p. 1582, 2022.

[52] M. Ge, N. F. Syed, X. Fu, Z. Baig, and A. Robles-Kelly, Towards a deep learning-driven intrusion detection approach for internet of things, *Computer Networks*, vol. 186, p. 107784, 2021.

[53] R. Malik, Y. Singh, Z. A. Sheikh, P. Anand, P. K. Singh, and T. C. Workneh, An improved deep belief network IDS on IoT-based network for traffic systems, *Journal of Advanced Transportation*, vol. 2022, p. 7892130, 2022.

[54] M. Alanazi and A. Aljuhani, Anomaly detection for internet of things cyberattacks, *Computers, Materials & Continua*, vol. 72, no. 1, pp. 261–279, 2022.
[55] J. D. Lee, H. S. Cha, S. Rathore, and J. H. Park, M-IDM: A multi-classification based intrusion detection model in healthcare IoT, *Computers, Materials & Continua*, vol. 67, no. 2, pp. 1537–1553, 2021.
[56] Z. K. Maseer, R. Yusof, S. A. Mostafa, N. Bahaman, O. Musa, and B. A. S. Al-rimy, DeepIoT.IDS: Hybrid deep learning for enhancing IoT network intrusion detection, *Computers, Materials & Continua*, vol. 69, no. 3, pp. 3945–3966, 2021.
[57] H. Yao, P. Gao, P. Zhang, J. Wang, C. Jiang, and L. Lu, Hybrid intrusion detection system for edge-based IIoT relying on machine-learning-aided detection, *IEEE Network*, vol. 33, no. 5, pp. 75–81, 2019.
[58] T. Kuang, Z. Hu, and M. Xu, A genetic optimization algorithm based on adaptive dimensionality reduction, *Mathematical Problems in Engineering*, vol. 2020, p. 8598543, 2020.
[59] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-IoT dataset, *Future Generation Computer Systems*, vol. 100, pp. 779–796, 2019.
[60] Q. Yang, J. Singh, and J. Lee, Isolation-based feature selection for unsupervised outlier detection, presented at Annu. Conf. Progn. Health Manag. Soc., Scottsdale, AZ, USA, 2019.
[61] M. Sarhan, S. Layeghy, and M. Portmann, Towards a standard feature set for network intrusion detection system datasets, *Mobile Networks and Applications*, vol. 27, no. 1, pp. 357–370, 2022.
[62] P. Nimbalkar and D. Kshirsagar, Feature selection for intrusion detection system in internet-of-things (IoT), *ICT Express*, vol. 7, no. 2, pp. 177–181, 2021.
[63] M. Abushwereb, M. Alkasassbeh, M. Almseidin, and M. Mustafa, An accurate IoT intrusion detection framework using apache spark, arXiv preprint arXiv: 2203.04347, 2022.
[64] T. Saba, A. Rehman, T. Sadad, H. Kolivand, and S. A. Bahaj, Anomaly-based intrusion detection system for IoT networks through deep learning model, *Computers & Electrical Engineering*, vol. 99, p. 107810, 2022.

Mouaad Mohy-Eddine received the master degree from Sultan Moulay Slimane University, Khouribga, Morocco in 2020. He is currently pursuing the PhD degree of computer security at Cadi Ayyad University, Marrakech. His main field of research interest is machine learning, intrusion detection, and IoT security.



Azidine Guezzaz received the PhD degree from Ibn Zohr University Agadir, Morocco in 2018. He is currently an assistant professor of computer science and mathematics at Cadi Ayyad University. His main field of research interest is computer security, cryptography, artificial intelligence, intrusion detection, and smart cities. He is also a reviewer of various scientific journals.

Mourade Azrour received the PhD degree from Moulay Ismail University of Meknès, Errachidia, Morocco, and the MS degree in computer and distributed systems from Ibn Zouhr University, Agadir, Morocco in 2014. He currently works as a compter sciences professor at the Faculty of Sciences and Techniques, Moulay Ismail University of Meknès. His research interests include authentication protocol, computer security, Internet of Things, smart systems. He is a member of the scientific committee of numerous international conferences. He is also a reviewer of various scientific journals. He has edited a scientific book *IoT and Smart Devices for Sustainable Environment* and he is a guest editor in journal *EAI Endorsed Transactions on Internet of Things*.

Said Benkirane received the PhD degree from Choaib Dokkali University, El jadida, Morocco in 2013. He is currently an associate professor of computer science and mathematics at Cadi Ayyad University, Marrakech. His research interests include computer security, artificial intelligence, smart cities, and VANET networks. He is also a reviewer of various scientific journals.

Yousef Farhaoui received the PhD degree in computer security from Ibn Zohr University of Science. He is a professor at the Faculty of Sciences and Techniques, Moulay Ismail University of Meknès, Morocco, and a local publishing and research coordinator, Cambridge International Academics in United Kingdom. His research interests include learning, e-learning, computer security, big data analytics, and business intelligence. He has three books in computer science. He is a coordinator and member of the organizing committee, a member of the scientific committee of several international congresses, and a member of various international associations. He has authored 4 books and many book chapters with reputed publishers such as Springer and IGI. He serves as a reviewer for IEEE, IET, Springer, Inderscience, and Elsevier journals. He is also the guest editor of many journals with Wiley, Springer, Inderscience, etc. He has been the general chair, session chair, and panelist in several conferences.