USING MACHINE LEARNING TO BUILD A CLASSIFICATION MODEL FOR IOT NETWORKS TO DETECT ATTACK SIGNATURES

Mousa Al-Akhras[1, 2], Mohammed Alawairdhi[1],
Ali Alkoudari[1] and Samer Atawneh[1]

[1]College of Computing and Informatics, Saudi Electronic University,
Riyadh 11673, Saudi Arabia
[2]Computer Information Systems Department, King Abdullah II School for Information Technology, The University of Jordan, Amman 11942, Jordan

ABSTRACT

Internet of things (IoT) has led to several security threats and challenges within society. Regardless of the benefits that it has brought with it to the society, IoT could compromise the security and privacy of individuals and companies at various levels. Denial of Service (DoS) and Distributed DoS (DDoS) attacks, among others, are the most common attack types that face the IoT networks. To counter such attacks, companies should implement an efficient classification/detection model, which is not an easy task. This paper proposes a classification model to examine the effectiveness of several machine-learning algorithms, namely, Random Forest (RF), k-Nearest Neighbors (KNN), and Naïve Bayes. The machine learning algorithms are used to detect attacks on the UNSW-NB15 benchmark dataset. The UNSW-NB15 contains normal network traffic and malicious traffic instants. The experimental results reveal that RF and KNN classifiers give the best performance with an accuracy of 100% (without noise injection) and 99% (with 10% noise filtering), while the Naïve Bayes classifier gives the worst performance with an accuracy of 95.35% and 82.77 without noise and with 10% noise, respectively. Other evaluation matrices, such as precision and recall, also show the effectiveness of RF and KNN classifiers over Naïve Bayes.

KEYWORDS

Internet of Things, Security, Classification model, Machine learning, Random Forest, k-Nearest Neighbors, Naïve Bayes.

1. INTRODUCTION

Internet of Things (IoT) is a network of devices that allows these devices to share information directed towards different purposes [1]. Such devices include desktops, laptops, smartphones, and tablets. The inception of smart devices to the society was first done in 1982, where the first device to ever be connected to the Internet was a Coca-Cola Company vending machine. This machine kept stock of its commodities and kept inventory for the inputs and outputs. The machine also monitored the temperature of the drinks within the machine. The term IoT was coined by Kevin Ashton of Proctor and Gamble in 1999. However, the actual existence of aspects that verified this term came into existence in 2008. Figure 1 shows the concept of IoT [2].

DOI: 10.5121/ijcnc.2020.12607

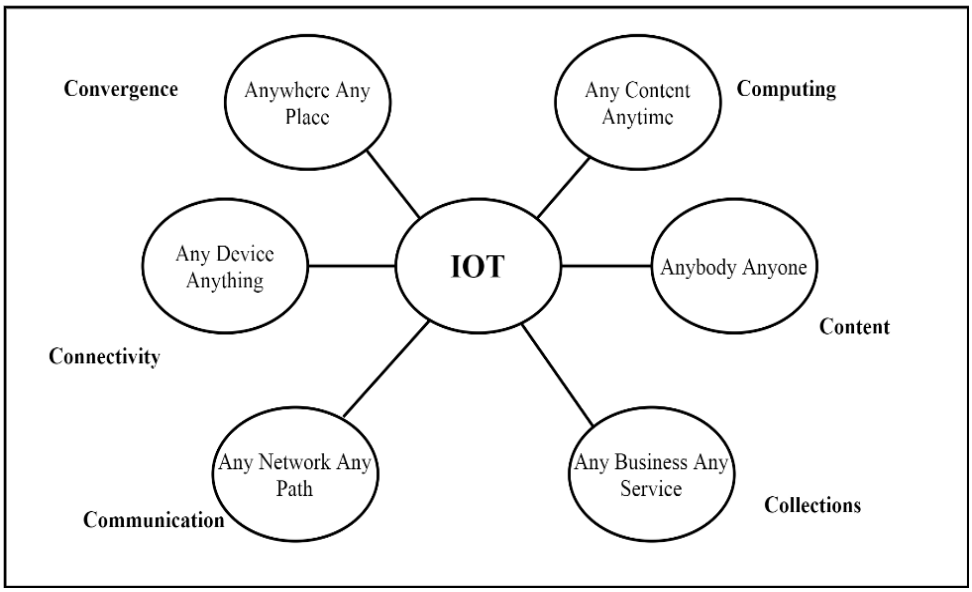

Figure 1. The concept of IoT

IoT applications are distributed across many fields including consumers, commercial, industrial, and infrastructure sectors. For the consumer field, the IoT is integrated into homes with aspects such as the existence of smart homes, that is, the homes with the ability to perform most of the essential functions that initially required human intervention. The abilities include the temperature regulation and security systems and actions such as fire prevention using smoke detectors. Additionally, IoT has been integrated into the healthcare systems with the use of mechanized patient data storage and analysis. Machines that can analyze an individual's health symptoms and facilitate the identification of diseases are also included in the healthcare sector as a result of the interaction of technology and the system. In the industrial sector, the use of IoT is embedded in the manufacturing processes. The use of manual labor has been reduced, and the efficiency of the processes enhanced as machines have taken up the processing, packaging, and sealing of products. Moreover, the use of computers in industrial processes had facilitated the achievement of higher levels of production as compared to when most of the factory processes were manually handled.

The use of IoT is also incorporated in Agriculture under the industrial bracket. Agriculture has been automated by the use of IoT with the likes of automated irrigation systems as well as the existence of climate-controlled greenhouses that make it possible to grow just about anything anywhere in the world [3]. On infrastructure, the use of IoT has been incorporated into aspects such as energy management. Through IoT, energy consumption can be regulated. The IoT is also used to conduct greenhouse environmental control and monitoring [4]. However, the continued interaction with the IoT has led to the predisposition of various challenges within society. Regardless the benefits that it has brought with it to the society, IoT have compromised security and privacy of individuals at various levels. Figure 2 shows the IoT security challenges [5].

Machine learning (ML) provides the programs with the ability to improve their performance with experience [6]. ML algorithms can be classified as supervised, unsupervised, and reinforcement learning. These categories can be used in areas weather forecasting, cluster identification and learning from mistakes, respectively. ML can assist the IoT arena by utilizing information on different security issues experienced and using it to make permanent solutions to the security threats for future preparedness. Also ML can help in detecting rare events or observations, i.e.,

the anomalies. Anomalies can raise suspicion because they are statistically different from other normal observations.

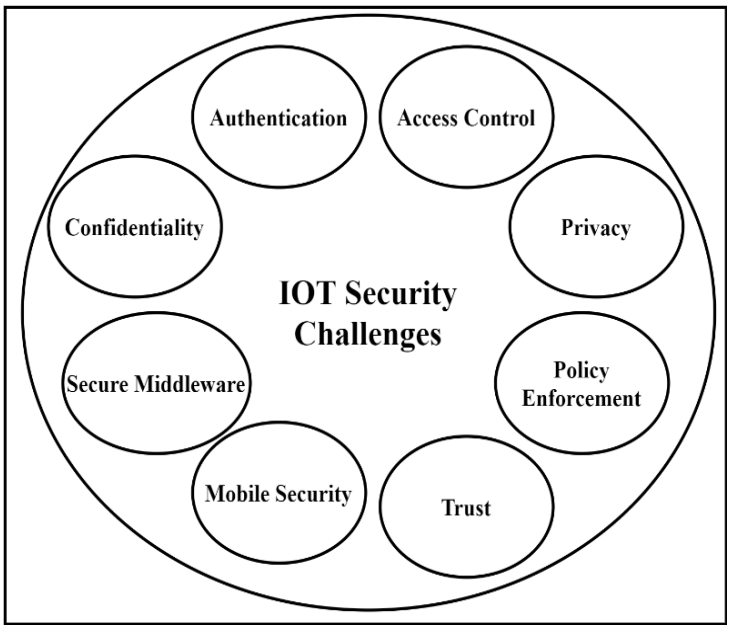

Figure 2. IoT Security Challenges

The contribution of this work is detecting anomalies using machine-learning model. Aclassification model is built to examine the effectiveness of a set ofML classification algorithms, namely, Random Forest (RF), k-Nearest Neighbors (KNN) algorithm, and Naïve Bayes, on a benchmark IoT dataset, known as UNSW-NB15, and estimate the appropriate selection of such algorithms for detecting anomalies in the IoT environment. A voting method will also be applied to improve the estimation process.

The rest of this paper is organized as follows. Section 2 presents the background and literature review. The details of the proposed classification model are given in Section 3 where different ML classifiers are applied to build the model. The model implementation results and discussion are presented in Section 4. Section 5 concludes the paper and suggests some possibilities of future works.

2. BACKGROUND AND LITERATURE REVIEW

IoT is a modern form of networking that oversees the conjoined performance of devices under one network. It involves the use of a network connection to run devices without the necessary intervention of human input or programming. Through IoT, there is the inception of smart cars in society, the automation of homes, the running of industrial processes among other functions that no longer require manual input. However, IoT is faced by several challenges alongside the benefits that it has brought with it to the society. The use of IoT is prone to several security issues that facilitate the corruption of its systems by individuals or organizations with malicious intent. This is because the systems involved in IoT are vulnerable to attacks due to some systematic loopholes that make it hard to contain these insecurities.

2.1. The Classification of IoT

Alaba et al. [3] proposed an IoT approach for the integration of various sensors and objects that can communicate with each other without the necessity of human intervention. IoT is described as inclusive of sensory devices that monitors and gathers all types of information on both machine and human social life. According to Hameed et al. [7], IoT is as a result of the interconnection of devices and networks as a result of the technological growth that has taken place since the last century. The anticipation of the increased interaction will lead to data generation at very high rates.

IoT is divided into three layers: the application layer, the perception layer, and the network layer. The application layer in IoT is the uppermost layer, which is visible to the end- users of IoT devices. The perception layer is the layer that is tasked with the collection of information. It includes the perception nodes as well as the perception network. Finally, the network layer provides network transmission and a pervasive access environment to the perception layer. According to Pishva [8], digital information has become a social infrastructure in our society. The perception of IoT is because of the interconnection of devices to one similar network, which is the Internet. The connectivity to the Internet is the biggest attraction for IoT. Japanese technologists spearheaded this with the inception of Internet-enabled audiovisual devices as early as ten years ago [8]. The integration of IoT has led to the establishment of the web economy. This is mainly because the Internet broke the physical barriers that limited trade to a regional basis only.

2.2. Security Issues and their Predisposing Factors in IoT

The complexity of IoT has resulted in the magnification of the security challenges that had initially been associated with the Internet. Node access, which is the basic functionality of the IoT, is a predisposing factor to the challenges that face this entity. A significant factor that contributed to the security challenges associated with IoT is the lack of software that protects devices on the IoT from viruses and malware attacks. Resource constraint solutions that mitigate the prevalence of attacks and enhance privacy protection are only included in traditional networks. Alaba et al. highlighted the security issues of IoT alongside the discussion of the conventional wireless network. The traditional wireless network is the traditional form of Internet that predisposed the coming of IoT. One major highlight is that IoT is based on the use of Low power and Lossy Networks (LLNs), which exposes it to data loss due to node impersonation. Security features differ between IoT and conventional networks. This is because sensor nodes have low computation power and low storage capacity that is a somewhat limiting factor considering the kind of data traffic IoT has to deal with. IoT also faces security issues such as false and man-in-the-middle attacks [3]. Both of these issues can capture information from the network and send fake data to the nodes in the network.

IoT lacks unified standards that could be the basis of the prevention of security challenges. The possible attacks identified include hardware and network threats. The future directions of IoT involve the address of heterogeneity, which is a nature of IoT that predisposes security risks that are associated with it[3]. Privacy is one of the security issues associated with IoT. The association of an individual's identity in IoT leads to the profiling and tracking of users. Malicious individuals can track users of IoT and profile their interactions with their environment due to the amount of information generated by IoT. To safeguard users, a secure method of data transmission should be included in the security of IoT to reduce the chances of an individual's invasion of privacy no matter the intent. Another predisposing factor for the insecurity affiliated with IoT is the lightweight cryptographic framework. IoT should be fixed to consume fewer

resources without the compromise on security. Currently, IoT consumes fewer resources but at the expense of safety which endangers the user [7].

Frustaci et al. [9] grouped the threats that are related to IoT based on the IoT Layer, which are perception, transportation, and application. Based on perception, which includes the physical aspects of IoT such as sensors and nodes that carry out data collection and perception, the threats include physical attack, impersonation, Denial of Service (DoS), routing attacks, and data transit attacks. Physical attacks involve the physical damage to hardware such as node tampering or the injection of malware directly into the system. On impersonation, it consists of the generation of fake identification through the use of malicious nodes. DoS attack consumes the finite network resources to prevent legitimate users from accessing the network. Routing attacks on the other hand focus on modifying the data routes during data collection and the forwarding process. A data transit attack involves attacks such as man-in-the-middle, which intercepts data and manipulates it according to the well of the hacker.

The transportation layer transmits the gathered information for the network. The security issues associated with the Transportation layer include routing attacks, DoS, and data transit attacks [9]. For the Application layer, the security issues include DoS, data leakage, and malicious code injection. Data leakage is the stealing of data based on its vulnerability. Frustaci et al. [9] highlighted the properties of trust in IoT and its importance which includes the certainty in collaboration, excellence in flexibility as well as the efficiency of the IoT. Trust within the IoT has led to the lack of use by Several Some Money organizations and individuals that feel that their data is too vulnerable. Through the enhancement of trust, IoT can develop to become globally adopted and used. While IoT is closed and customers cannot add security software's to the devices, traditional networks have the option of adding antiviruses and other security controls [9]. IoT can also only manage to use lightweight algorithms, which cause the high affinity to attacks as it is focused on balancing higher security with the low capacity of devices. Traditional IT was user-controlled while IoT automatically collects private information about the user. The devices of traditional IT are located in closed environments while those in the IoT are located in open environments.

Xiao et al. [6] noted that IoT has facilitated the integration between the physical world and communication networks and its application in the environment we live in. The study highlighted the need to address of security issues affiliated with IoT such as spoofing, intrusion, Distributed DoS (DDoS), jamming, eavesdropping, and malware.

2.3. Current Solutions to the Security Issues Facing IoT

To enhance the security situation associated with IoT, there should be an enhancement of its features to serve this purpose. One way is through the establishment of secure routes for the sharing of data and information all around the globe. Another preventive measure is the isolation of malicious nodes, which are used by hackers for their malice acts. The system should be redesignedto enable the detection and isolation of malicious nodes. The system should also be enhanced to perform self-stabilization following an attack [7]. The security protocol should be able to guide the network to recover without the need for human intervention. To enhance security, the preservation of location privacy should be included in the system. In regards to robustness and resilience, the network should be designed to tolerate intrusion and malware attacks by detecting the attackers as early as possible.

It should also promote quick recovery from failures that may come as a result of the attack. Thesystem should be self-reliant and does not require the intervention of humans in order to recover or protect the users from attack.Hameed et al. [7] investigated DoS attacks which bar

individuals from their services within the networks as part of a cyber-attack aimed at gaining some information from the culprit. This calls for an efficient resource counter measure alongside the implementation of a resource efficient insider attack detection. These properties can be usedas prevention measures to prevent DoS and DDoS. The current challenges that IoT encounters are discussed in [7].

Pishva [8] highlighted the vulnerabilities that are affiliated with Internet connectivity. Some of the risks associated with the Internet include the leakage of information, privacy infringement, and data corruption. Some of these risks are created by Internet service providers but to enhance the service provision. However, these loopholes created are abused for business expansion at theexpense of user victimization as well as the use by attackers to perform their malicious acts. Another aspect is the abuse of privacy in E-commerce. Buyers currently interact with goods and service providers online. However, this has become a problem for individuals on the Internet as the marketers utilize individual's information and bombard them with junk mail based on their purchases. One factor that facilitates attacks on the users of IoT is the lack of cryptographic capabilities. This is because the higher number of devices connected to the IoT is our primary function devices that are cheap and do not ascertain the move of using expensive software for their protection. The existence of technology unaware individuals is also a factor that has led to the situation that predisposes insecurity in IoT. For traditional users, the device would only be connected to the Internet when it is in use and it will be switched off when there is no interaction with it while smart devices are always connected which facilitates the attacks.

One proposed security measure is the use of the United Home Gateway (UHG). This single pathway connects all household devices equipped with appropriate security measures. It prevents the access of the devices, as it is much easier to compromise them when they are directly connected to the Internet. Other proposed solutions include changing defaults passwords, disconnection of universal Plug and Play (PnP) features as they created security loopholes for an IoT device. The last proposition includes keeping the software up to date as this usually fixes security loopholes and bugs. By keeping a device's software updated, an individual stands a higher chance of protecting their devices from attack.

2.4. Machine Learning in Developing IoT Security

In order to enhance the safety of use for IoT, methods such as authentication, access control, malware detections, and secure off loading are highlighted as the best security measures for IoT security. Authentication helps IoT devices to distinguish the source nodes and address the identity-based attacks such as Sybil and spoofing attacks [6]. This prevents the interaction with malicious nodes that could predispose the cases of identification attacks. Another remedy to the security issues in the IoT is the use of an access control mechanism. Access control prevents the access of IoT resources by unauthorized users. Thus, IoT can only be used by limited legitimate individuals who are allowed to access the resources on a specific device. In addition, the inclusion of secure offloading techniques enables IoT devices to utilize the storage and computation of IoT servers.

As mentioned in Section 1, ML provides machines the ability to learn and improve its performance in accordance with the previous experience without the necessary input of a person. The ML can assist the IoT arena by utilizing information on the different security issues experienced and using it to make permanent solutions to the security threats for future preparedness. This section reviews current articles that utilize ML algorithms to enhance the security aspects of an IoT network. Andročec and Vrček conducted several studies on the types of ML algorithms to tackle the security of the IoT [10]. The authors noted that high number IoTdevices have compromised the security of systems. They stated that numerous cases of

malicious software that attacks and damages Internet devices and systems. This has led to the use of ML algorithms to promote IoT security [10]. In this section, the authors of the current paper extended the work of Andročec and Vrček [10] by carrying out a thorough review of recent research papers. Thirty-four studies conducted recently were identified and analyzed to show the importance of using ML in tackling the security issues of IoT. The findings revealed that more studies on ML for IoT security had become mature.

Table 1 summarizes the reviews of the papers that use ML algorithms, besides other techniques, in the IoT security. For each study, the attack types, the used security techniques, the ML techniques, a brief summary, the type of dataset used, and whether the dataset is noisy or not as shown in Table 1. The reviewed papers show the importance of using the ML algorithms to promote the security of the IoT and hence improving the attacks detection process, however, to the best of our knowledge, no paper studies the importance of applying a voting algorithm in detecting the attacks. The voting algorithm combines different ML algorithms in order to improve the prediction results and thus enhancing the security of IoT.

Table 1. Reviews of papers that use ML algorithms in the IoT security.

Ref.	Attack types	Securitytechniques	ML techniques	Summary	Dataset used	Noise
[11]	Code confiscation	Dynamic monitoring	Naive Bayes	Discussing the concepts of code confiscation as a strategy that can help to create a secure mechanism for guaranteeing a secure architecture.	Constructed data	No
[12]	Cyber attacks	Efficient behavior approach	Averaged One-Dependence Estimator (AODE)	Advancements in technology brought more challenges to the IoT field. Efficient capturebehavior acts as a solution to the difficulties witnessed.	Constructed data	No
[13]	Cyber attacks	Blockchain approach	Reinforcement Learning (RL) algorithm	Dynamic access control policy is described in details and its abilities to provide the ultimate solution to security issues. The work uses the blockchain strategy and machine learning to create a solution.	Constructed dataset	No
[14]	Cyber-physical attacks	Authentication	Extreme learning machine (ELM)	ML is the right approach for ensuring the ultimate security of devices.	Consntruced Data	Yes
[15]	Cyber-physical attacks	Adaptive authentication	Naive Bayes	eHealth faces numerous challenges today. Overcoming these challenges may be rectified by the use of a risk-based adaptive authentication approach.	UNSW dataset	No
[16]	DDoS	DPMM	DPMM	ML is one of the most preferred strategies for guaranteeing the security of various networks.	Consntruced Data	No
[17]	DoS	Access control	Q-learning	Discussing the remotestate estimation of cyberphysical systems under signal-to-interference-plus-noise ratio-basedDoS attacks.	Consntruced Data	Yes
[18]	DoS	Access control	Multivariate correlationanalysis	DoS attack detection model based onMultivariate Correlation Analysis to characterizenetwork traffic accurately.	KDD cup 99	No
[19]	Hacking attacks	Sparse approximation	Approximate Nearest Neighbor	Sparse approximation technique plays a critical role in the creation of the ultimate solution to approach the security of IoT.	Consntruced Data	Yes

[20]	Hacking attacks	adaptive biometric and authentication	Support Vector Machine using Gaussian radial basis function (SVM,GF)	Security of the IoT remains a significant challenge. The work proposes a use of an authentication system as the ultimate solution to the security issues witnessed.	Consntruced Data	No
[21]	Internet traffic analysis	Differentiated privacy framework	machine learning	Smart homes in the wake of IoT are targets by attackers. The detection and guaranteeing of security to such aspects remains high.	Consntruced Data	No
[22]	Intrusion	Use IoT IDM	Support Vector Machines (SVMs)	Securing Home IoT devices remains a challenge. However, the work claims that using OpenFlow; users can create a secure platform to achieve the desired outcomes.	IoT Profile	No
[23]	Intrusion	Deep abstraction	SVM andArtificial Neural Networks (ANN)	Wifi impersonation is one of the latest attacks and threats that users face. However, wifi impersonation can be rectified by the use of deep abstraction and weighted feature selecting techniques.	AWID	Yes
[24]	Intrusion	Deep abstraction	Deep learning (DL)	Distributed attacks are rampant nowadays especially in the. DL plays an important role in creating an ultimate solution for such challenges.	NSL-KDD	No
[25]	Intrusion	Deep abstraction	Used classifiers: Logistic, Linear, SVM,Decision Tree Classifier (Gini, Entropy)	Creating a reliable strategy for achieving security of various devices in IoT is a challenge to many.	kdd cup 99	No
[26]	Intrusion	SDN	SVM	SDN is one of the most recent strategies that companies may use to secure their systems.	Constructed data	No
[27]	Intrusion	low footprint unsupervised learning	One-class SVM	Creating the ultimate security mechanism for the cloud is a challenge today. However, the work suggests that intelligent security is one of the best strategies to use when it comes to the creation of the ultimate solution.	AIS dataset	No
[28]	Intrusion	Intrusion detection analysis	ANN and SVM	The work claims that the security of computer networks require the adoption of effective approaches such as machine learning.	NSL-KDD dataset (noise)	Yes
[29]	Intrusion	Intrusion detection analysis	Decision-Tree Learning	UsieML classifiers play an important role in securing network devices.	Constructed data	No
[30]	Intrusion	Behavior Profiling	k-Means algorithm and SVM	Abnormal behavior profiling working together with machine learning may create an ultimate security solution to IoT architecture.	Constructed data	No
[31]	Intrusion	Intrusion detection	ANN	Implementing best intrusion detection strategies for determining the ultimate network security may become a challenge. Using radio communication profiling remains an ultimate solution to the underlying aspects.	Constructed data	No
[32]	Intrusion	Intrusion detection	ANN	ML is crucial in the determination of the best security aspects of IoT.	Constructed data	No
[33]	Intrusion	Intrusion detection	ANN and Decision tree	Securing WIFI connections remains a challenge to companies. Usingweighted feature selection may help to create a secure platform forguaranteeing the security of networks.	AWID	No

[34]	Intrusion	Intrusion detection	Random forest multi-way classified	ML acts as a solution to determining the right strategies to achieve data security. SecuringIoT devices requires ML approach and algorithms.	Constructed data	Yes
[35]	Intrusion	Distributed attack detection	Deep learning	Adoptinga new approach, deep learning, to cybersecurity toenable the detection of attacks in social IoT.	NSL-KDD dataset	No
[36]	Intrusion	Intrusion detection	Random forest	The work presents best strategies to use when it comes to the security of IoT. Some of the selected strategies include ML.	Constructed data	Yes
[37]	Intrusion	Access control	SVM Naïve Bayes	The work presents an extensive literature review over the period 2002–2013 of ML methods that were used to address common issues in WSNs.	Constructed data	No
[38]	Intrusion	Access control	ANN	The work describes the use of ML and data mining methods for cyber analytics in support of intrusion detection.	DARPA	Yes
[39]	Jamming	Secure IoT offloading	DQN	Dynamic anti-jamming communication game for CRNs, which exploits spreadspectrum and user mobility to improve SINR of the signalsagainst cooperative smart jammers	Constructed data	Yes
[40]	Spoofing	Authentication	Q-learning Dyna-Q	Formulating the interactions between a receiver and spoofers as a zero-sum spoofing detection game. A PHY authentication method based on Q-learning and Dyna-Q is usedfor dynamic radio environment.	Constructed data	Yes
[41]	Spoofing	Authentication	DNN	User authenticationwith a device-free approach by leveraging commonWiFi signals of IoT devices, such assmart TV and refrigerator.	Constructed data	Yes
[42]	Spoofing	Authentication	dFW	The work proposes a physical-layerauthentication system that exploits the channel state informationof radio transmitters to detect spoofing attacks in wirelessnetworks.	Constructed data	No
[43]	Frequency fingerprinting	Authentication	SVM, KNN, and Decision Trees	The work highlights the strategies to use to secure IoT using radio frequency fingerprint.	Constructed data	No
[44]	Hacking	Authentication	Dispersion Entropy (DE)	The creation of the ultimate security strategy for securing the IoT acts as an ultimate challenge. However, using the physical layer authentication approach may play a critical role in creating the ultimate solution.	Constructed data	No

Section 3 presents the contribution of this work by introducing the proposed classification model. The proposed model detects the anomalies and hence increases the protection of the IoT environment. This can be achieved by examining the effectiveness of a set of ML classification algorithms, namely, Random Forest (RF), k-Nearest Neighbors (KNN) algorithm, and Naïve Bayes, on the IoT dataset, known as UNSW-NB15, to estimate the appropriate selection of such algorithms for detecting anomalies. The proposed model also applies a voting algorithm to produce an efficient predictive model and hence improving the estimation process over other ML

classifiers. The details of the proposed classification model are given in Section 3 while its implementation results and discussion are presented in Section 4.

3. THE PROPOSED CLASSIFICATION MODEL

This section presents the details of the proposed classification model. Different ML classifiers are applied to build a model, which can classify legal and illegal traffics that are generated under different scenarios. Weka platform shown in Figure 3 has a collection of ML classifiers for data mining. Weka has been used in the implementation of the proposed model. Weka contains different tools for data regression, classification, clustering, visualization and association rules mining. The following subsections present the used measures, the used ML classifiers, and the benchmark dataset that will be used in the proposed model.

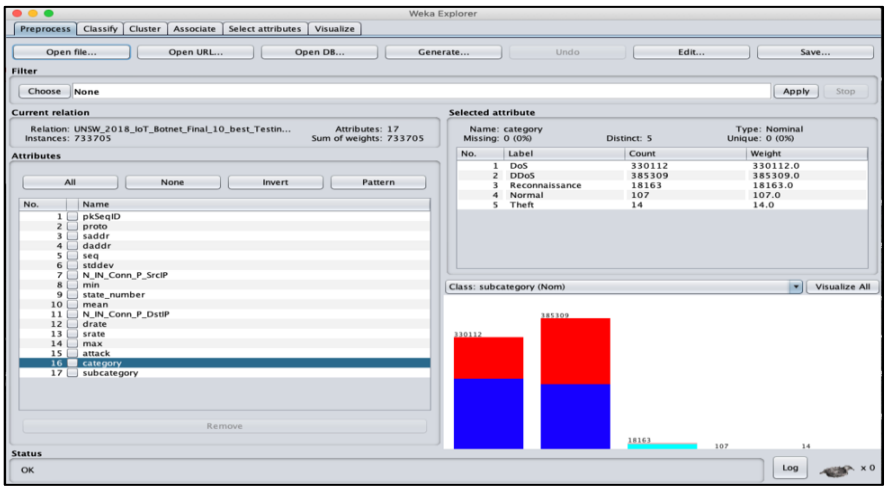

Figure 3.The Weka platform

3.1. Evaluation Metrics

Confusion metrics will be used to evaluate the performance of the classifiers. Confusion metrics are commonly used in classification problems that have two or more types of classes. The used confusion metrics in this research are accuracy, precision, and recall.

3.1.1. Accuracy

In classification problems, the accuracy is the number of correct predictions over all predictions made. It is a good measure when the target variable classes are nearly balanced.

3.1.2. Precision

Precision shows, out of all the positive observations, how many positive observations are predicted correctly. The higher number of correct predictions means the higher the performance of the classifier. Prediction can be given using the following equation:

$$\text{Precision}=TP/(TP+FP), \quad (1)$$

where TP represents true positive predictions and FP represents false positive predictions [45].

3.1.3. Recall

Recall shows, out of all observations in the actual class, how many positive observations are predicted correctly. As precision, the higher number of correct predictions is the higher the performance of the classifier. Recall can be given using the following equation:

$$\text{Recall} = TP/(TP+FN), \quad (2)$$

where FN represents the false negative prediction [45].

3.2. Classification Algorithms

In the proposed model, the classification algorithms (i.e., classifiers) that will be used are Random Forest (RF), K-Nearest Neighbors (KNN), and Naïve Bayes. In addition to these algorithms, a voting algorithm will be applied in the proposed model. A voting method is an ML classifier that combines different models in order to produce an optimal predictive model that leads to improve the prediction results. The following subsections briefly present the used classifiers.

3.2.1. Random Forest

Random Forest (RF) that is commonly usedbecause of its simplicity and the fact that it can be used for both regression and classification tasks. RF classifier creates several, but random, decision trees and merges them to produce a more accurate and stable prediction model [46].

3.2.2. K-Nearest Neighbours (KNN)

KNN algorithm supports both classification and regression. It stores a training dataset and conducts queries on the data set to locate the k most similar patterns to make predictions. KNN algorithm takes the category of the most similar items in the dataset and assign this category to the unlabeled instances [47].

3.2.3. Naïve Bayes

Naïve Bayes algorithm is a learning and a statistical method for classification. It is constructed using the training data set, and it estimates the probability of each class given the features of new instances [48].

The proposed classification model will be applied on the UNSW-NB15 dataset, which is a modern IoT dataset that covers a collection of a large number of normal network traffic and malicious traffic instances.UNSW-NB15 dataset is a benchmark dataset used to detect network's malicious activities [49]. The University of New South Wales created the UNSW-NB15 dataset for evaluating new Intrusion Detection Systems (IDS). One hundred GB of raw network traffic were collected to build UNSW-NB15 dataset. The dataset consists of 45 attributes and it has ten categories of traffic; one normal and nine categories represent different forms of attacks. Figure 4 illustrates the workflow of the proposed classification model.

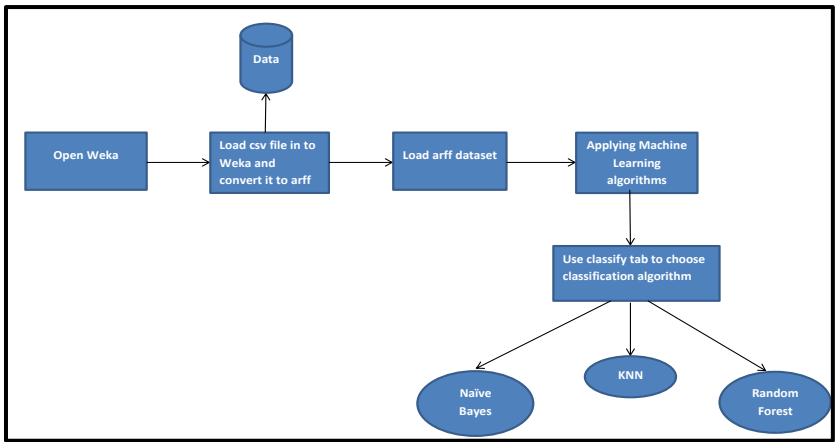

Figure 4. The workflow of the proposed classification model

4. RESULTS AND DISCUSSION

This section presents the results of the proposed classification model. Different ML classifiers, namelyKNN, Random Forest, Naïve Bayes and a voting method, were applied in Weka platform to get the results. The proposed model has been implemented using a large number of normal network traffic and malicious traffic instances taken from a modern IoT dataset known under the name UNSW-NB15. Different experiments, using set of classifiers, have been conducted as shown in Figures 5 – 12.

Based on the experimental results, the percentage of correctly classified instances was 100% for the RF and KNN classifiers, while the accuracy percentage for the Naïve Bayes classifier was about 95%. Furthermore, the experimental results reveal that the evaluation metrics accuracy, precision, and recall have the highest values for the RF and KNN classifiers. However, the same evaluation metrics have the lowest values for Naïve Bayes classifier (about 95% in average for each metric). In addition, after adding 10% of noise, the experimental results indicate that the RF classifier has the highest performance while the Naïve Bayes classifier still shows the worst performance. The findings also reveal that the proposed Voting algorithm has the best performance over the evaluation metrics accuracy, precision, and recall.

Total Number of Instances: 220111
Correctly Classified Instances: 220111 (100 %)
Incorrectly Classified Instances: 0 (0 %)
Mean absolute error: 0
Relative absolute error: 0.0078 %
Root mean squared error: 0.0016
Root relative squared error: 0.4813 %
Kappa statistic: 1
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	ROC Area	Class
	1.000	0.000	1.000	1.000	1.000	**DoS**
	1.000	0.000	1.000	1.000	1.000	**DDoS**
	1.000	0.000	1.000	1.000	1.000	**Reconnaissance**
	1.000	0.000	1.000	1.000	1.000	**Normal**
	1.000	0.000	1.000	1.000	1.000	**Theft**
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	

Figure 5. Random Forest Classifier results summary

Total Number of Instances: 220111
Correctly Classified Instances: 220110 (99.9995 %)
Incorrectly Classified Instances: 1 (0.0005 %)
Mean absolute error: 0.0001
Relative absolute error: 0.0248 %
Root mean squared error: 0.0028
Root relative squared error: 0.8754 %
Kappa statistic: 1
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	ROC Area	Class
	1.000	0.000	1.000	1.000	1.000	**DoS**
	1.000	0.000	1.000	1.000	1.000	**DDoS**
	1.000	0.000	1.000	1.000	1.000	**Reconnaissance**
	0.971	0.000	1.000	0.971	1.000	**Normal**
	1.000	0.000	1.000	1.000	1.000	**Theft**
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	

Figure 6. Random Forest Classifier (Noise) results summary

Total Number of Instances: 220111
Correctly Classified Instances: 220111 (100 %)
Incorrectly Classified Instances: 0 (0 %)
Mean absolute error: 0
Relative absolute error: 0.0015 %
Root mean squared error: 0
Root relative squared error: 0.0012 %
Kappa statistic: 1
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	ROC Area	Class
	1.000	0.000	1.000	1.000	1.000	**DoS**
	1.000	0.000	1.000	1.000	1.000	**DDoS**
	1.000	0.000	1.000	1.000	1.000	**Reconnaissance**
	1.000	0.000	1.000	1.000	1.000	**Normal**
	1.000	0.000	1.000	1.000	1.000	**Theft**
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	

Figure 7. KNN classifier results summary

Total Number of Instances: 220111
Correctly Classified Instances: 220090 (99.9905 %)
Incorrectly Classified Instances: 21 (0.0095 %)
Mean absolute error: 0
Relative absolute error: 0.0198 %
Root mean squared error: 0.0062
Root relative squared error:1.914 %
Kappa statistic: 0.9998
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	ROC Area	Class
	1.000	0.000	1.000	1.000	1.000	**DoS**
	1.000	0.000	1.000	1.000	1.000	**DDoS**
	0.998	0.000	1.000	0.998	0.999	**Reconnaissance**
	1.000	0.000	1.000	1.000	1.000	**Normal**
	1.000	0.000	0.500	1.000	1.000	**Theft**
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	

Figure 8. KNN Classifier (Noise) results summary

Total Number of Instances: 220111
Correctly Classified Instances: 209877 (95.3505 %)
Incorrectly Classified Instances: 10234 (4.6495 %)
Mean absolute error: 0.0171
Relative absolute error: 8.2129 %
Root mean squared error: 0.1177
Root relative squared error: 36.4597 %
Kappa statistic: 0.9104
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	ROC Area	Class
	0.920	0.019	0.976	0.920	0.994	**DoS**
	0.980	0.076	0.934	0.980	0.994	**DDoS**
	0.991	0.000	1.000	0.991	0.999	**Reconnaissance**
	1.000	0.000	1.000	1.000	1.000	**Normal**
	1.000	0.000	1.000	1.000	1.000	**Theft**
Weighted Avg.	0.954	0.048	0.955	0.954	0.994	

Figure 9. Naïve Bayes results summary

Total Number of Instances: 220111
Correctly Classified Instances: 209792 (95.3119 %)
Incorrectly Classified Instances: 10319 (4.6881 %)
Mean absolute error: 0.0173
Relative absolute error: 8.2856 %
Root mean squared error: 0.1183
Root relative squared error: 36.6393 %
Kappa statistic: 0.9096
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	ROC Area	Class
	0.919	0.019	0.976	0.919	0.994	**DoS**
	0.980	0.077	0.934	0.980	0.994	**DDoS**
	0.990	0.000	1.000	0.990	0.998	**Reconnaissance**
	0.941	0.000	1.000	0.941	1.000	**Normal**
	1.000	0.000	1.000	1.000	1.000	**Theft**
Weighted Avg.	0.953	0.049	0.954	0.953	0.997	

Figure 10. Naïve Bayes (Noise) results summary

Total Number of Instances: 220111
Correctly Classified Instances: 220111(100 %)
Incorrectly Classified Instances: 0 (0 %)
Mean absolute error: 0.0057
Relative absolute error: 2.7407 %
Root mean squared error: 0.0392
Root relative squared error: 12.1563 %
Kappa statistic: 1
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	ROC Area	Class
	1.000	0.000	1.000	1.000	1.000	**DoS**
	1.000	0.000	1.000	1.000	1.000	**DDoS**
	1.000	0.000	1.000	1.000	1.000	**Reconnaissance**
	1.000	0.000	1.000	1.000	1.000	**Normal**
	1.000	0.000	1.000	1.000	1.000	**Theft**
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	

Figure 11. Voting Algorithm results summary

Total Number of Instances: 220111
Correctly Classified Instances: 220107 (99.9982 %)
Incorrectly Classified Instances: 4 (0.0018 %)
Mean absolute error: 0.0058
Relative absolute error: 2.7762 %
Root mean squared error: 0.0395
Root relative squared error: 12.2452 %
Kappa statistic: 1
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	ROC Area	Class
	1.000	0.000	1.000	1.000	1.000	**DoS**
	1.000	0.000	1.000	1.000	1.000	**DDoS**
	1.000	0.000	1.000	1.000	1.000	**Reconnaissance**
	1.000	0.000	1.000	1.000	1.000	**Normal**
	1.000	0.000	1.000	1.000	1.000	**Theft**
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	

Figure 12. Voting Algorithm (Noise) results summary

Table 2 compares the results of several ML algorithms on the UNSW-NB15 dataset. Experimental results show that the RF, KNN and Voting algorithms are the best performers with an accuracy of 100%, while Naïve Bayes classifier was the worst performer with an accuracy of 95.35%. Furthermore, after applying 10% of noise to the data, the experimental results indicate that the RF classifier has the highest performance. The results also reveal that the voting algorithm has the best performance over the evaluation metrics accuracy, precision, and recall.

Table 2. Comparison Results between the applied ML classifiers

ML algorithm	Accuracy	Precision	Recall
KNN	100	1	1
KNN (Noise)	99.9905	1	1
RF	100	1	1
RF (Noise)	99.9995	1	1
Naïve Bayes	95.3505	0.955	0.954
Naïve Bayes (Noise)	82.7719	0.820	0.828
Vote	100	1	1
Vote (Noise)	99.9982	1	1

5. CONCLUSIONS AND FUTURE WORK

IoT networks have several security and privacy challenges. To address these challenges, this workdeveloped a classification model to examine the effectiveness of well-known machine learning algorithms, namely, RF, KNN and Naïve Bayes, in detecting attack signatures (i.e., anomalies) on a benchmark dataset that contains both normal network traffic and malicious traffic instances. A set of evaluation matrices, namely, accuracy, precision and recall, were used in evaluating the performance of the classifiers. Both RF and KNN classifiers showed much better performance (in all evaluation metrics) than the Naïve Bayes classifier. Moreover, to improve the prediction results, a voting method that combines several base models was implemented in the proposed classifier model and it revealed the best performance over all evaluation metrics. This work is limited to study the performance of only three machine-learning algorithms in detecting attack signatures. However, the work can be expanded in future by studying the performance of other machine learning classifiers using different evaluation metrics on different datasets.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Gupta, P.; Agrawal, D.; Chhabra, J.; Dhir, P.K. In *Iot based smart healthcare kit*, 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT), 2016; IEEE: pp 237-242.
2. Mustafa, G.; Ashraf, R.; Mirza, M.A.; Jamil, A. In *A review of data security and cryptographic techniques in iot based devices*, Proceedings of the 2nd International Conference on Future Networks and Distributed Systems, 2018; pp 1-9.
3. Alaba, F.A.; Othman, M.; Hashem, I.A.T.; Alotaibi, F. Internet of things security: A survey. *Journal of Network and Computer Applications* **2017**, *88*, 10-28.
4. Wang, J.; Chen, M.; Zhou, J.; Li, P. Data communication mechanism for greenhouse environment monitoring and control: An agent-based iot system. *Information Processing in Agriculture* **2019**.
5. Sicari, S.; Rizzardi, A.; Grieco, L.A.; Coen-Porisini, A. Security, privacy and trust in internet of things: The road ahead. *Computer networks* **2015**, *76*, 146-164.
6. Xiao, L.; Wan, X.; Lu, X.; Zhang, Y.; Wu, D. Iot security techniques based on machine learning. *arXiv preprint arXiv:1801.06275* **2018**, 1-20.
7. Hameed, S.; Khan, F.I.; Hameed, B. Understanding security requirements and challenges in internet of things (iot): A review. *Journal of Computer Networks and Communications* **2019**, *2019*, 1-14.
8. Pishva, D. Iot: Their conveniences, security challenges and possible solutions. *Adv. Sci. Technol. Eng. Syst. J* **2017**, *2*, 1211-1217.
9. Frustaci, M.; Pace, P.; Aloi, G.; Fortino, G. Evaluating critical security issues of the iot world: Present and future challenges. *IEEE Internet of Things Journal* **2017**, *5*, 2483-2495.
10. Andročec, D.; Vrček, N. In *Machine learning for the internet of things security: A systematic review*, The 13th International Conference on Software Technologies, 2018.
11. Cho, T.; Kim, H.; Yi, J.H. Security assessment of code obfuscation based on dynamic monitoring in android things. *Ieee Access* **2017**, *5*, 6361-6371.
12. Ali, T.; Nauman, M.; Jan, S. Trust in iot: Dynamic remote attestation through efficient behavior capture. *Cluster Computing* **2018**, *21*, 409-421.
13. Outchakoucht, A.; Hamza, E.; Leroy, J.P. Dynamic access control policy based on blockchain and machine learning for the internet of things. *Int. J. Adv. Comput. Sci. Appl* **2017**, *8*, 417-424.
14. Wang, Z.; Chen, Y.; Patil, A.; Jayabalan, J.; Zhang, X.; Chang, C.-H.; Basu, A. Current mirror array: A novel circuit topology for combining physical unclonable function and machine learning. *IEEE Transactions on Circuits and Systems I: Regular Papers* **2017**, *65*, 1314-1326.
15. Gebrie, M.T.; Abie, H. In *Risk-based adaptive authentication for internet of things in smart home ehealth*, Proceedings of the 11th European Conference on Software Architecture: Companion Proceedings, 2017; ACM: pp 102-108.
16. Ahmed, M.E.; Kim, H.; Park, M. In *Mitigating dns query-based ddos attacks with machine learning on software-defined networking*, MILCOM 2017-2017 IEEE Military Communications Conference (MILCOM), 2017; IEEE: pp 11-16.
17. Li, Y.; Quevedo, D.E.; Dey, S.; Shi, L. Sinr-based dos attack on remote state estimation: A game-theoretic approach. *IEEE Transactions on Control of Network Systems* **2016**, *4*, 632-642.
18. Tan, Z.; Jamdagni, A.; He, X.; Nanda, P.; Liu, R.P. A system for denial-of-service attack detection based on multivariate correlation analysis. *IEEE transactions on parallel and distributed systems* **2013**, *25*, 447-456.
19. Razeghi, B.; Voloshynovskiy, S.; Kostadinov, D.; Taran, O. In *Privacy preserving identification using sparse approximation with ambiguization*, 2017 IEEE Workshop on Information Forensics and Security (WIFS), 2017; IEEE: pp 1-6.
20. Yeh, K.-H.; Su, C.; Hsu, C.-L.; Chiu, W.; Hsueh, Y.-F. In *Transparent authentication scheme with adaptive biometrie features for iot networks*, 2016 IEEE 5th Global Conference on Consumer Electronics, 2016; IEEE: pp 1-2.
21. Liu, J.; Zhang, C.; Fang, Y. Epic: A differential privacy framework to defend smart homes against internet traffic analysis. *IEEE Internet of Things Journal* **2018**, *5*, 1206-1217.

22. Nobakht, M.; Sivaraman, V.; Boreli, R. In *A host-based intrusion detection and mitigation framework for smart home iot using openflow*, 2016 11th International conference on availability, reliability and security (ARES), 2016; IEEE: pp 147-156.
23. Aminanto, M.E.; Choi, R.; Tanuwidjaja, H.C.; Yoo, P.D.; Kim, K. Deep abstraction and weighted feature selection for wi-fi impersonation detection. *IEEE Transactions on Information Forensics and Security* **2017**, *13*, 621-636.
24. Abeshu, A.; Chilamkurti, N. Deep learning: The frontier for distributed attack detection in fog-to-things computing. *IEEE Communications Magazine* **2018**, *56*, 169-175.
25. Indre, I.; Lemnaru, C. In *Detection and prevention system against cyber attacks and botnet malware for information systems and internet of things*, 2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP), 2016; IEEE: pp 175-182.
26. Bhunia, S.S.; Gurusamy, M. In *Dynamic attack detection and mitigation in iot using sdn*, 2017 27th International Telecommunication Networks and Applications Conference (ITNAC), 2017; IEEE: pp 1-6.
27. Zissis, D. In *Intelligent security on the edge of the cloud*, 2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC), 2017; IEEE: pp 1066-1070.
28. Perez, D.; Astor, M.A.; Abreu, D.P.; Scalise, E. In *Intrusion detection in computer networks using hybrid machine learning techniques*, 2017 XLIII Latin American Computer Conference (CLEI), 2017; IEEE: pp 1-10.
29. Gao, S.; Thamilarasu, G. In *Machine-learning classifiers for security in connected medical devices*, 2017 26th International Conference on Computer Communication and Networks (ICCCN), 2017; IEEE: pp 1-5.
30. Lee, S.-Y.; Wi, S.-r.; Seo, E.; Jung, J.-K.; Chung, T.-M. In *Profiot: Abnormal behavior profiling (abp) of iot devices based on a machine learning approach*, 2017 27th International Telecommunication Networks and Applications Conference (ITNAC), Australia, 2017; IEEE: Australia, pp 1-6.
31. Roux, J.; Alata, E.; Auriol, G.; Nicomette, V.; Kaâniche, M. In *Toward an intrusion detection approach for iot based on radio communications profiling*, 2017 13th European Dependable Computing Conference (EDCC), 2017; IEEE: pp 147-150.
32. Canedo, J.; Skjellum, A. In *Using machine learning to secure iot systems*, 2016 14th Annual Conference on Privacy, Security and Trust (PST), 2016; IEEE: pp 219-222.
33. Aminanto, M.E.; Tanuwidjaja, H.C.; Yoo, P.D.; Kim, K. In *Wi-fi intrusion detection using weighted-feature selection for neural networks classifier*, 2017 International Workshop on Big Data and Information Security (IWBIS), 2017; IEEE: pp 99-104.
34. Wu, M.; Song, Z.; Moon, Y.B. Detecting cyber-physical attacks in cybermanufacturing systems with machine learning methods. *Journal of intelligent manufacturing* **2019**, *30*, 1111-1123.
35. Diro, A.A.; Chilamkurti, N. Distributed attack detection scheme using deep learning approach for internet of things. *Future Generation Computer Systems* **2018**, *82*, 761-768.
36. Domb, M.; Bonchek-Dokow, E.; Leshem, G. Lightweight adaptive random-forest for iot rule generation and execution. *Journal of Information Security and Applications* **2017**, *34*, 218-224.
37. Alsheikh, M.A.; Lin, S.; Niyato, D.; Tan, H.-P. Machine learning in wireless sensor networks: Algorithms, strategies, and applications. *IEEE Communications Surveys & Tutorials* **2014**, *16*, 1996-2018.
38. Buczak, A.L.; Guven, E. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials* **2015**, *18*, 1153-1176.
39. Han, G.; Xiao, L.; Poor, H.V. In *Two-dimensional anti-jamming communication based on deep reinforcement learning*, 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017; IEEE: pp 2087-2091.
40. Xiao, L.; Li, Y.; Han, G.; Liu, G.; Zhuang, W. Phy-layer spoofing detection with reinforcement learning in wireless networks. *IEEE Transactions on Vehicular Technology* **2016**, *65*, 10037-10047.
41. Shi, C.; Liu, J.; Liu, H.; Chen, Y. In *Smart user authentication through actuation of daily activities leveraging wifi-enabled iot*, Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing, 2017; ACM: pp 1-10.
42. Xiao, L.; Wan, X.; Han, Z. Phy-layer authentication with multiple landmarks with reduced overhead. *IEEE Transactions on Wireless Communications* **2017**, *17*, 1676-1687.

43. Baldini, G.; Steri, G.; Giuliani, R.; Gentile, C. In *Imaging time series for internet of things radio frequency fingerprinting*, 2017 International Carnahan Conference on Security Technology (ICCST), 2017; IEEE: pp 1-6.
44. Baldini, G.; Giuliani, R.; Steri, G.; Neisse, R. In *Physical layer authentication of internet of things wireless devices through permutation and dispersion entropy*, 2017 Global Internet of Things Summit (GIoTS), 2017; IEEE: pp 1-6.
45. Al-issa, A.I.; Al-Akhras, M.; ALsahli, M.S.; Alawairdhi, M. In *Using machine learning to detect dos attacks in wireless sensor networks*, 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), 2019; IEEE: pp 107-112.
46. Alizadeh, Z.; Mohammadizadeh, M. Predicting electron-phonon coupling constants of superconducting elements by machine learning. *Physica C: Superconductivity and its Applications* **2019**, *558*, 7-11.
47. Zhang, Z. Introduction to machine learning: K-nearest neighbors. *Annals of translational medicine* **2016**, *4*.
48. Guzmán-Cabrera, R.; Sánchez, B.P.; Mukhopadhyay, T.P.; García, J.; Córdova-Fraga, T. Classification of opinions in cross domains involving emotive values. *Journal of Intelligent & Fuzzy Systems* **2019**, *36*, 4877-4887.
49. Moustafa, N.; Slay, J. In *Unsw-nb15: A comprehensive data set for network intrusion detection systems (unsw-nb15 network data set)*, 2015 military communications and information systems conference (MilCIS), 2015; IEEE: pp 1-6.