BIG DATA MINING AND ANALYTICS
ISSN 2096-0654 05/10 pp311–320
Volume 6, Number 3, September 2023
DOI: 10.26599/BDMA.2022.9020038

Cloud-Based Intrusion Detection Approach Using Machine Learning Techniques

Hanaa Attou, Azidine Guezzaz*, Said Benkirane, Mourade Azrour, and Yousef Farhaoui

Abstract: Cloud computing (CC) is a novel technology that has made it easier to access network and computer resources on demand such as storage and data management services. In addition, it aims to strengthen systems and make them useful. Regardless of these advantages, cloud providers suffer from many security limits. Particularly, the security of resources and services represents a real challenge for cloud technologies. For this reason, a set of solutions have been implemented to improve cloud security by monitoring resources, services, and networks, then detect attacks. Actually, intrusion detection system (IDS) is an enhanced mechanism used to control traffic within networks and detect abnormal activities. This paper presents a cloud-based intrusion detection model based on random forest (RF) and feature engineering. Specifically, the RF classifier is obtained and integrated to enhance accuracy (ACC) of the proposed detection model. The proposed model approach has been evaluated and validated on two datasets and gives 98.3% ACC and 99.99% ACC using Bot-IoT and NSL-KDD datasets, respectively. Consequently, the obtained results present good performances in terms of ACC, precision, and recall when compared to the recent related works.

Key words: cloud security; anomaly detection; features engineering; random forest

1 Introduction

Cloud technologies allow practical access on demand to a shared network, storage, and resources and offer more choices regarding their service models[1]. These models are platform as a service (PaaS), software as a service (SaaS), and infrastructure as a service (IaaS)[2], used in one of the deployment models private, public, and hybrid cloud[3]. The cloud provides services with high performance due to its characteristics[2] according to the National Institute of Standards and Technology[4]: network access, resource pooling, quick elasticity, and measured service. Recently, the cloud suffers from many security problems like availability, data confidentiality, integrity, and control authorization. In addition, the Internet is used to facilitate access to the services offered by the cloud representing a major source of threats that can infect the cloud systems and resources[2]. Then enhancing cloud security becomes a primary challenge for cloud providers[5]. Therefore, several approaches such as firewall tools, data encryption algorithms, authentication protocols, and others have been developed to better secure cloud environments from various attacks[6]. However, traditional systems are not sufficient to secure cloud services from different limits[7]. Therefore, a set of intrusion detection approaches are proposed and applied to detect and prevent undesirable activities in real-time[8,9]. In general, the detection methods are divided into misuse detection method which uses known attacks to detect intrusion and anomaly detection method which detect intrusion using unknown attack. The hybrid

• Hanaa Attou, Azidine Guezzaz, and Said Benkirane are with the Technology Higher School Essaouira, Cadi Ayyad University, Marrakech 44000, Morocco. E-mail: A.GUZZAZ@gmail.com.
• Mourade Azrour and Yousef Farhaoui are with the STI Laboratory, the IDMS team, Faculty of Sciences and Techniques, Moulay Ismail University of Meknès, Errachidia 25003, Morocco.
* To whom correspondence should be addressed.
Manuscript received: 2022-09-07; revised: 2022-09-27; accepted: 2022-10-12

method is obtained by combining the advantages of these two methods[10]. Despite of more solutions given to secure cloud environments, the recent intrusion detection systems (IDSs) are affected by various significant limitations[8], for example, huge amounts of analyzed data, real-time detection, data quality, and others that aim to decrease the performance of detection models. Nowadays, academic researchers show that intelligent learning methods[6, 11] such as machine learning (ML), deep learning (DL), and ensemble learning are useful in various areas[12, 13] and are able to perform network security[14–18]. Our main goal in this research work is to propose an anomaly detection approach based on random forest (RF) binary classifier and feature engineering is carried out based on a data visualization process aiming to reduce the number of used features and perform the proposed anomaly detection model. The evaluation performances of the model are implemented on NSL-KDD and BoT-IoT datasets. Then, the obtained outcomes demonstrate model performances. The rest of this paper is described as follows. In Section 2, we present the state-of-the-art cloud computing (CC) architectures, IDS, ML methods, and recent related works in the domain. The proposed framework is presented in Section 3. In Section 4, we demonstrate the experimental Setting. Then, in Section 5, we describe in detail the obtained results. The paper is achieved with a conclusion and future works.

2 State-of-the-Art Works

In this section, we present a state-of-the-art CC architecture, IDS, ML methods, and related works that describe different algorithms used to enhance IDS and cloud security.

The models of cloud services are IaaS, PaaS, and SaaS. They differ according to the technical layers offered[19]. The IaaS model provides temporary virtual machines (VMs) and also allows an increase in the storage space of VMs, networks, and load balancers. They offer the technical layers of IaaS in addition to middleware instances and execution contexts, such as databases and application servers, whereas PaaS models only offer middleware instances. They are provided over the Internet on-demand and with a measured service[20]. SaaS models offer software[20] and users could run desired applications as shown in Fig. 1.

Cloud deployment models are intended for different entities as needed[2]. The public cloud is a model that intended its resources for public clients, as the name suggests. However, the private cloud is only for one entity. The hybrid cloud concept combines both private and public clouds. Community cloud is a multi-tenant platform that allows multiple companies to collaborate on the same platform if their needs and concerns are similar[2]. The most important difference between the public cloud and the private cloud is that the private cloud is considered the most secure since it has fewer users than the public cloud[2, 5]. Intrusion is a kind of unauthorized activity that could pose a possible threat to the information's confidentiality, integrity, and availability[8]. Researchers have developed IDS that aims to detect any type of intrusion. It achieves this objective by monitoring activities at the network or host machine. Depending on these activities, IDSs consist of two basic varieties, Network IDS and Host IDS[8, 21]. We can distinguish between misuse based detection and anomaly based detection. The first one is used to detect known attacks and the second one is

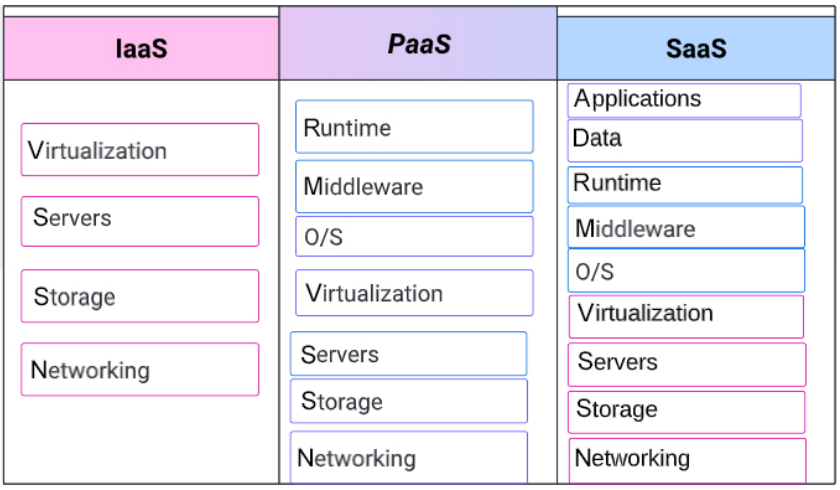


Fig. 1 Cloud computing models.

used to detect unknown attacks. Hybrid based detection combines both methods for the ability to detect known and unknown attacks[12]. ML is a technique used to feed the IDS for identifying attacks[22]. Computers may learn without being explicitly programmed thanks to the ML field[23]. Concretely, it is a branch of modern science that uses statistics to identify patterns in data and then make predictions[24]. We can subdivide ML into three types of learning[25]. To discover a relationship from a series of samples, a supervised ML technique is utilized. The learned association can then be used to forecast data that have not been seen before. The most known supervised learning algorithms are k-nearest neighbors (KNNs), decision trees (DTs), linear regression (LR), neural networks (NNs), and RF[26]. Ferrag et al.[27] used the DT algorithms classifier to train the IDS in a layered approach. Unsupervised learning is unlike supervised learning, data are unlabeled and machine learns without an example[25]. Some of the important unsupervised learning are clustering, visualization, dimensionality reduction, and association rule learning[23]. Semi-supervised learning is between supervised and non-supervised learning[23]. Besides, deep learning (DL) is a type of ML method based on learning data representations. Also it is the main ML technology that relies on algorithms of artificial neural networks (ANNs)[27]. On the other sides, firewalls are used by cloud providers to identify intrusions, but this technology does not detect insider attacks[7]. Therefore, the challenge is to detect the different types of intrusions in the cloud. There are different related studies that use DL and ML techniques to reinforce computer security. Kanimozhi and Jacob[28] used a calibration curve to evaluate the different classifier methods such as KNN classifier, Naïve Bayes (NB), Adaboost with DT, support vector machine (SVM) classifier, and RF classifier to detect a portrayal of botnet attacks on dataset CSE-CIC-IDS2018. Zhou et al.[29] suggested a deep neural network (DNN) based IDS. In particular, the system employs three phases: data acquisition (DAQ), data pre-processing, and then DNN classifications. With SVM, the system obtains an accuracy (ACC) of 0.963. A software-defined networking IDS using a DL method was put forth by Tang et al.[30] A two-stage DL technique for intrusion detection that focuses on finding malicious assaults on autonomous cars was developed by Zhang et al.[31] They started with a reliable rule-based system and then switched to a DL system in the second stage to detect anomalies. In Ref. [32], Mishra et al. proposed a classification based ML approach to detect distributed denial of service (DDoS) attacks in CC using three methods (KNN, RF, and NB), the proposed model achieves 99.76% ACC, and they concluded that RF gives the best results. Alshammari and Aldribi[22] applied ML techniques to feed IDS and detect malicious network traffic in cloud computing, and ISOT-CID is the dataset used to evaluate the performance. Jiang et al.[33] tested the effectiveness of the suggested attack detection system using the NSL-KDD dataset and concluded that long short term memory (LSTM) and recurrent neural networks (RNNs) are the best choices for multichannel IDS. The system's efficiency is reported to be 99.23% and its ACC to be 98.94%. To learn from privacy preserved encrypted data on cloud, Khan et al.[34] used supervised and unsupervised ML specifically ANNs over the scrambled information. SNORT and optimized back propagation neural network (BPN) have been proposed as a cooperative and hybrid network intrusion detection framework by Chiba et al.[7] This system attempts to improve BPN algorithm by merging signature-based detection (SNORT) with anomaly-based detection BPN. In Ref. [35], the model used is DL to build two classes, since the database used NSL KDD contains 39 types of attacks that are grouped into 4 classes, this study showed that working only on two classes normal or anomaly. Kim et al.[36] suggested an architecture for intrusion detection that uses the LSTM as a recurrent neural network and the KDD Cup 99 dataset. As an input vector, they employed 41 features. In Ref. [37], Zhang proposed an automatic technique that develops the discriminative model and fuses multi-view information to improve accuracy (ACC). Six basic features are used by Tang et al.[30] to build an IDS based on DL. According to the attack detection performance, the suggested system obtains an ACC of 96.93%. In Ref. [38], Ahmad et al. proposed a method for cloud-based text document classification and data integrity. Ahmad et al.[38] concluded that RF outperforms the different techniques used NB, SVM, and KNN. Recently, Mubarakali et al.[39] used SVM-based expert systems to detect distributed denial of service (DDoS). The system performance is reported as 96.23%. The comparison of various current IDS models is shown in Table 1.

3 Proposed Framework

In this section, we described various techniques used to elaborate our solution. The features reducing are

Table 1 Comparison of various current IDS models.

Reference	Year	Method used	ACC (%)	Dataset
Chiba et al.[7]	2016	BPN	–	–
Alshammari and Aldribi[22]	2021	ANN	92.00	ISOT-CID
		KNN	100.00	
		DT	100.00	
		SVM	81.00	
		NB	60.00	
		RF	100.00	
Kanimozhi and Jacob[28]	2019	ANN	99.90	CSE-CIC-IDS 2018
		RF	99.80	
		KNN	99.73	
		SVN	99.80	
		Adaboost	99.90	
		NB	99.20	
Zhou et al.[29]	2018	DNN	96.30	–
Tang et al.[30]	2016	DNN	75.75	NSL-KDD
Zhang et al.[31]	2018	DNN	–	–
Mishra et al.[32]	2021	RF, KNN, NB	99.76	–
Jiang et al.[33]	2018	LSTM	98.94	NSL-KDD
Khan et al.[34]	2019	ANNs	–	–
Potluri and Diedrich[35]	2016	DL	97.50	NSL-KDD
Kim et al.[36]	2016	LSTM	96.93	KDD CUP'99
Ahmad et al.[38]	2022	RF, NB, SVM, KNN	92.00	–
Mubarakali et al.[39]	2020	SVM	96.23	–

integrated to minimize execution time and to perform prediction. In addition, the RF classifier is trained with the two features selected as a subset from the NSL-KDD dataset to identify intrusions. As illustrated in Fig. 2, our proposed model adopts the main standard components of IDS. Hence, our model implements the collecting data module, preprocessing module data, and decision module. Our contribution focused more on the preprocessing data module by enhancing feature engineering tasks and obtaining reliable predictions.

The preprocessing module focused on data normalization. Therefore, the categorical features are transformed into numeric values with the dummies function that allows symbolic features to be mapped as numeric values. Then, the detected inconsistencies are deleted. Our intrusion detection model includes feature selection to identify and combine useful features for accurate detection. The graphic data visualization task is used to select the optimum feature subset that can enhance the prediction of the proposed model. Once the subset is selected, the RF algorithm is applied to obtain a reliable classifier to distinguish between

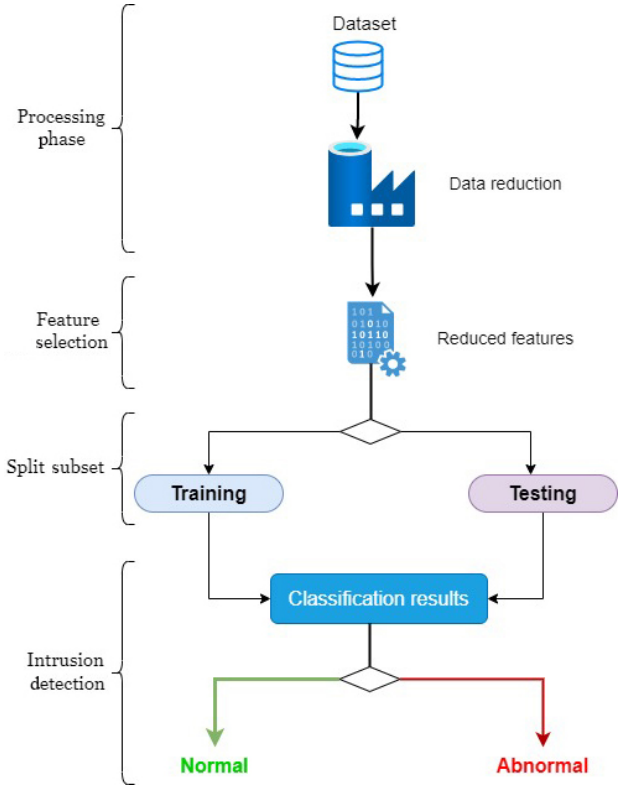


Fig. 2 Proposed model architecture.

normal or abnormal activities. The RF algorithm groups many DTs whose outputs merged into one final output[40]. DT is a supervised learning algorithm without a hyper-parameter designed for classification and regression[40]. Breiman[41] mentioned the idea that RF performs well unlike other classifiers like SVM, neural networks, and discriminant analysis. It avoids the over-adjustment issue. DT tends to overfit, whereas RF uses a bagging method to deal with these issues[42].

- RF as defined in Refs. [40, 43, 44] employs a classifier combination.
- Base classifiers using an m tree structure $\{h(X, \Theta n), N = 1, 2, \ldots, m\}$.
- X represents the input data, and $\{\Theta n\}$ is a dependent distributed random vector.
- Every DT selects data randomly from the available data.
- Build a forest to build the number of trees "n" by repeating the steps above for the number of times "n".

According to Refs. [42, 45], the advantages of RF are

- They are less sensitive to outlier data due to their ability to overcome the issue of overfitting in training data. It is simple to establish parameters, which avoids the requirement for tree trimming. Variable importance and ACC are automatically created.

• RF is a classifier that includes a group of tree-structured learners that each cast a majority election for the class that received the most input. Using the training test, a tree is constructed and is independent of earlier random vectors of the same distribution, and an upper bound is derived for RF to get the prediction error in terms of two factors that are exactitude and interdependence of different classifiers.

4 Experimental Setting

Our research work is carried out and evaluated in an experimental setting using a computer with a Core TM-i5 8250U CPU running at 1.8 GHz and 12 GB of RAM running windows 10 professional 64 bits. While Python 3 is used to implement the RF, DT, and SVM models after graphic visualization is used to reduce features. To validate our proposed model, we evaluate the ACC metric and compare it with the ACC of other models. As a result, we divided the entire dataset randomly. In the training step, 70% are employed, and the last part is used in the test step. The best parameters for any classifier performance are determined by the dataset used in the model training and testing. In this research work, NSL-KDD and Bot-IoT datasets are used. In order to address some of the inherent issues with the KDD 1999 dataset, a new edition of the KDD dataset was developed[46]. The NSL-KDD dataset provides the following advantages over the original KDD dataset: It excludes records that are redundant or duplicated. The amount of records is appropriate and selected records are arranged as a percentage of records (80% eKDDTrain+20% ARFF). The forty-one initial features from the KDD'99 dataset are available in NSL-KDD[46]. The NSL-KDD dataset yields the best results. Hence, 41 features are included in our dataset. The NSL-KDD dataset's six fundamental properties are utilized to develop various models, as stated in Ref. [30].

• **Duration:** Duration of the connection in seconds.

• **protocol_type:** Protocol_type_tcp, protocol_type_udp, and protocol_type_icmp are the three different types of protocols.

• **src_bytes:** Data bytes sent from the source to the destination.

• **dst_bytes:** Number of data bytes sent between source and destination.

• **Count:** Number of connections made to the same host in the previous two seconds as the connection type.

• **srv_count:** Number of prior two-second connections to the same service as the current connection.

The protocol_type categorical variable was converted into numeric values using the dummies function. According to graphic visualization shown in Fig. 3, we conclude that the class variable is not influenced by protocol_types variable.

The class_anomaly variable can be predicted from the variables count, duration and dst_host_srv_count only in a few points as shown in Fig. 4. For example, if the duration variable >1500 then we can detect an anomaly.

Figure 5a shows that we can predict class variable if src_bytes variable > 0. Also, we can detect that class_anomaly equals 0 if the variable dst_bytes is higher than 50 000 as shown in Fig. 5b.

The selected variables from visualization are src_bytes and dst_bytes, and then we reduce the number of features from 41 to two features. As the first step, after the graphic visualization, RF model was developed for class_anomaly and the selected variables from graphic visualization src_bytes and dst_bytes to detect intrusion. The Bot-IoT dataset is more developed since it includes IoT devices that work with both simulated and actual data[27,47]. Shafiq et al.[48] identified the top five variables with improved characteristics using ML approaches including DT, NB, RF, and SVM as well as measures like Pearson moment correlation and area under the curve (AUC). This dataset contains information on numerous distinct forms of IoT traffic flows, including regular traffic, IoT traffic, and botnet

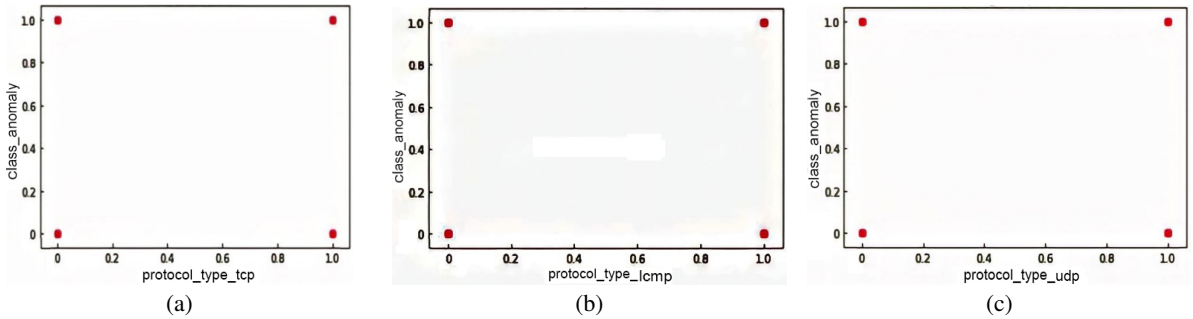
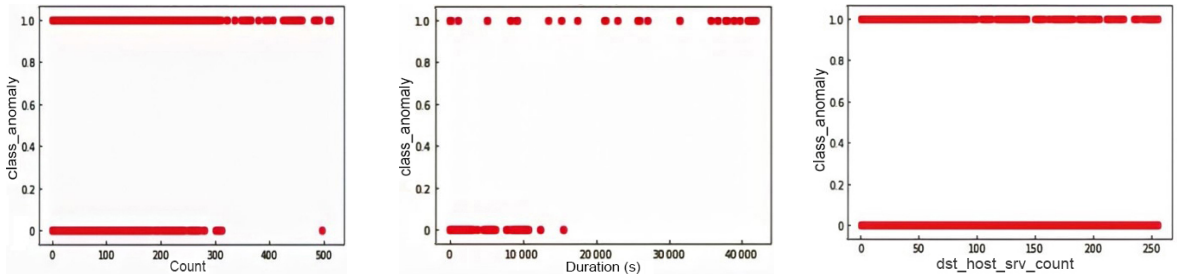
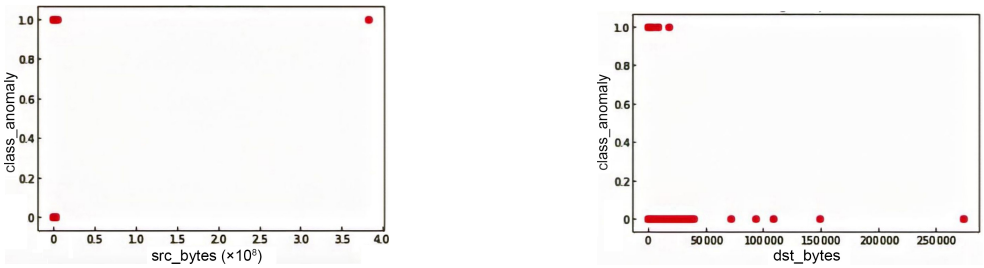


Fig. 3 Graphic visualization for different variables of protocol_type.



(a) (b) (c)

Fig. 4 Graphic visualization.



(a) (b)

Fig. 5 Graphic visualization of the selected variables.

traffic[49].

In order to evaluate our proposed model, we have selected randomly two features:

- **State_number:** Feature state is represented numerically.
- **Stddev:** The standard deviation of records that have been aggregated.

5 Result and Discussion

5.1 Evaluation metrics

We focus on classification models and precisely binary classification since intrusion detection is a problem when we use labeled data to predict whether the object belongs to the attack class or not. The results given by a binary classification algorithm are 0 or 1. Choosing the right metric is therefore crucial for evaluating and validating ML models. In these types of problems, the metrics generally consist of comparing the actual classes to the classes predicted by the model. This makes it possible to interpret the predicted probabilities for these classes. The key performance metric for classification is the confusion matrix[50] which is visualization, in table form, of the predictions of the model in relation to the real labels. The instances of a real class are represented in each row and those of a predicted class are represented in each column of the confusion matrix. From this matrix, we can calculate the different metrics ACC, recall, and precision which will allow us to evaluate our IDS based on RF.

- ACC is calculated using Eq. (1). It is the percentage of accurate predictions made relative to all cases.

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \tag{1}$$

- Recall is obtained from Eq. (2). It is used to determine the percentage of correctly categorized positive patterns.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{TN}} \times 100\% \tag{2}$$

- Precision is calculated using Eq. (3). It is out of all the expected patterns. A precision measurement counts the number of accurately predicted positive patterns in a positive class.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \tag{3}$$

5.2 Obtained results

Initially, the IDS is implemented for the classification task. The ACC value influences how well the model performs. The RF classification model was first improved by identifying the features that produce the best classification outcomes. After that, we used a subset of the NSL-KDD dataset to train the model.

Hence, we have started the evaluation of our model based on the two selected variables from the NSL-KDD src_bytes and dst_bytes using graphic visualization. To prove that the chosen variables are efficient, we use matrix correlation which is a table showing the correlation values for several variables. In Fig. 6 the

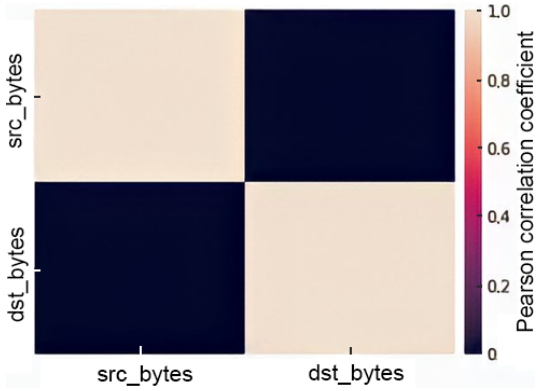


Fig. 6 Pearson correlation.

matrix depicts the correlation between src_bytes and dst_bytes, we conclude from Fig. 6 that there is no relation between the two variables since the coefficient tends toward 0 which shows that the risk of the multicollinearity is negligible. The outcomes illustrated in Fig. 7 show that our model performs well in terms of ACC and precision using src_bytes and dst_bytes, but the recall requires an improvement.

On the other hand, to check the effectiveness of the model, BoT-IoT is employed. We aggregate the data into a new data frame after importing it separately. Then we follow the steps outlined in our model displayed in Fig. 2. Two features are selected from BoT-IoT: state_number and stddev. Once we have done our tests, the results are established in Fig. 7. In Fig. 7, we display the obtained results of the three metrics that are used to evaluate the performance and efficiency of our proposed model. As it is very clear, we have obtained 98.3% of ACC, 96.3% of precision, and 46.0% of recall using NSL-KDD dataset. Furthermore, both ACC, precision, and recall reached 100% when the BoT-IoT dataset is used.

Figure 8 shows the ACC obtained by the different models using NSL-KDD and the ACC of our model using NSL-KDD and BoT-IoT. Our proposed IDS performs well if we compare it with the works proposed in Refs. [30, 33, 35, 39]. We obtained a higher ACC

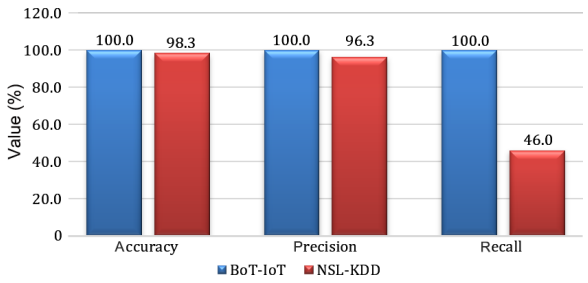


Fig. 7 Performance metrics of our model with NSL-KDD and BoT-IoT datasets.

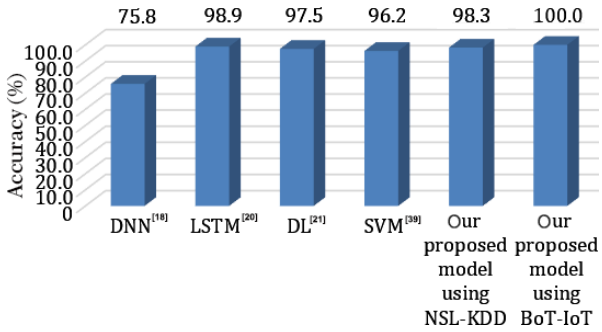


Fig. 8 Performance comparison between the proposed model and other works.

with the use of two selected features from NSL-KDD and BoT-IoT and RF than the other works mentioned in Fig. 8.

Consequently, reducing number of explanatory variables reduces data collection time and execution time. Hence, we maintain high quality results as shown in Fig. 8. As a result of all of this, we have demonstrated that our RF classifier technique can distinguish between normal and aberrant traffic using only two features. Besides, the RF gives good results compared to DNN, LSTM, DL, and SVM.

6 Conclusion

Intrusion detection is a new technology that has improved the security of the cloud. Recently, ML algorithms have been used to develop this technique because they are very helpful to secure and monitor systems. In this paper, we present an approach for detecting intrusions by combining graphic visualization and RF for cloud security. Then the first one is used for features engineering and the second one is used to predict and detect intrusions. Before the training of the model, we reduced the number of features to two. Based on the obtained results, the RF classifier is a remarkably more accurate method to predict and classify the attack type than DNN, DT, and SVM. We have demonstrated the potential of using a small number of features by contrasting the results with those of other classifiers. But recall is still not well enough using NSL-KDD, so in future work, we will focus on this point by using DL and ensemble learning techniques to improve our model.

References

[1] M. Ali, S. U. Khan, and A. V. Vasilakos, Security in cloud computing: Opportunities and challenges, *Information Sciences*, vol. 35, pp. 357–383, 2015.

[2] A. Singh and K. Chatterjee, Cloud security issues and challenges: A survey, *Journal of Network and Computer*

Applications, vol. 79, pp. 88–115, 2017.

[3] P. S. Gowr and N. Kumar, Cloud computing security: A survey, *International Journal of Engineering and Technology*, vol. 7, no. 2, pp. 355–357, 2018.

[4] A. Verma and S. Kaushal, Cloud computing security issues and challenges: A survey, in *Proc. First International Conference on Advances in Computing and Communications*, Kochi, India, 2011, pp. 445–454.

[5] H. Alloussi, F. Laila, and A. Sekkaki, L'état de l'art de la sécurité dans le cloud computing: Problèmes et solutions de la sécurité en cloud computing, presented at Workshop on Innovation and New Trends in Information Systems, Mohamadia, Maroc, 2012.

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

[7] Z. Chiba, N. Abghour, K. Moussaid, A. E. Omri, and M. Rida, A cooperative and hybrid network intrusion detection framework in cloud computing based snort and optimized back propagation neural network, *Procedia Computer Science*, vol. 83, pp. 1200–1206, 2016.

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

[9] A. Guezzaz, A. Asimi, Y. Asimi, Z. Tbatou, and Y. Sadqi, A global intrusion detection system using PcapSockS sniffer and multilayer perceptron classifier, *International Journal of Network Security*, vol. 21, no. 3, pp. 438–450, 2019.

[10] 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.

[11] B. A. Tama and K. H. Rhee, HFSTE: Hybrid feature selections and tree-based classifiers ensemble for intrusion detection system, *IEICE Trans. Inf. Syst.*, vol. E100.D, no. 8, pp. 1729–1737, 2017.

[12] 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, pp. 2793–2801, 2022.

[13] M. Azrour, Y. Farhaoui, M. Ouanan, and A. Guezzaz, SPIT detection in telephony over IP using K-means algorithm, *Procedia Computer Science*, vol. 148, pp. 542–551, 2019.

[14] M. Azrour, M. Ouanan, Y. Farhaoui, and A. Guezzaz, Security analysis of Ye et al. authentication protocol for internet of things, in *Proc. International Conference on Big Data and Smart Digital Environment*, Casablanca, Morocco, 2018, pp. 67–74.

[15] 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.

[16] 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.

[17] A. Guezzaz, A. Asimi, M. Azrour, Z. Tbatou, and Y. Asimi, A multilayer perceptron classifier for monitoring network traffic, in *Proc. 3rd International Conference on Big Data and Networks Technologies*, Leuven, Belgium, 2019, pp. 262–270.

[18] S. Benkirane, Road safety against sybil attacks based on RSU collaboration in VANET environment, in *Proc. 5th International Conference on Mobile, Secure, and Programmable Networking*, Mohammedia, Morocco, 2019, pp. 163–172.

[19] Q. Zhang, L. Cheng, and R. Boutaba, Cloud computing: State-of-the-art and research challenges, *J. Internet Serv. Appl.*, vol. 1, pp. 7–18, 2010.

[20] M. K. Srinivasan, K. Sarukesi, P. Rodrigues, M. S. Manoj, and P. Revathy, State-of-the-art cloud computing security taxonomies: A classification of security challenges in the present cloud computing environment, in *Proc. 2012 International Conference on Advances in Computing, Communications and Informatics*, Chennai, India, 2012, pp. 470–476.

[21] 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.

[22] A. Alshammari and A. Aldribi, Apply machine learning techniques to detect malicious network traffic in cloud computing, *Journal of Big Data*, vol. 8, p. 90, 2021.

[23] A. Géron, *Hands-On Machine Learning with Scikit-Learn & TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Sebastopol, CA, USA: O'Reilly Media, Inc., 2017.

[24] N. Chand, P. Mishra, C. R. Krishna, E. S. Pilli, and M. C. Govil, A comparative analysis of SVM and its stacking with other classification algorithm for intrusion detection, in *Proc. 2016 International Conference on Advances in Computing, Communication, & Automation* (*ICACCA*), Dehradun, India, 2016, pp. 1–6.

[25] A. B. Nassif, M. A. Talib, Q. Nasir, H. Albadani, and F. M. Dakalbab, Machine learning for cloud security: A systematic review, *IEEE Access*, vol. 9, pp. 20717–20735, 2021.

[26] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, and K. J. Kim, A survey of deep learning-based network anomaly detection, *Cluster Comput.*, vol. 22, pp. 949–961, 2017.

[27] 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.

[28] V. Kanimozhi and T. P. Jacob, Calibration of various optimized machine learning classifiers in network intrusion detection system on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing, *International Journal of E*ngineering Applied Sciences and Technology, vol. 4, no. 6, pp. 209–213, 2019.

[29] L. Zhou, X. Ouyang, H. Ying, L. Han, Y. Cheng, and T. Zhang, Cyber-attack classification in smart grid via deep neural network, in *Proc. 2nd International Conference on Computer Science and Application Engineering*, Hohhot,

China, 2018, pp. 1–5.

[30] T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, and M. Ghogho, Deep learning approach for network intrusion detection in software defined networking, in *Proc. 2016 International Conference on Wireless Networks and Mobile Communications*, Fez, Morocco, 2016, pp. 258–263.

[31] L. Zhang, L. Shi, N. Kaja, and D. Ma, A two-stage deep learning approach for can intrusion detection, in *Proc. 2018 Ground Vehicle Syst. Eng. Technol. Symp.* (*GVSETS*), Novi, MI, USA, 2018, pp. 1–11.

[32] A. Mishra, B. B. Gupta, D. Perakovic, F. J. G. Penalvo, and C. H. Hsu, Classification based machine learning for detection of DDoS attack in cloud computing, in *Proc. 2021 IEEE International Conference on Consumer Electronics*, Las Vegas, NV, USA, 2021, pp. 1–4.

[33] F. Jiang, Y. Fu, B. B. Gupta, Y. Liang, S. Rho, F. Lou, F. Meng, and Z. Tian, Deep learning based multi-channel intelligent attack detection for data security, *IEEE Transactions on Sustainable Computing*, vol. 5, no. 2, pp. 204–212, 2018.

[34] A. N. Khan, M. Y. Fan, A. Malik, and R. A. Memon, Learning from privacy preserved encrypted data on cloud through supervised and unsupervised machine learning, in *Proc. 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies*, Sukkur, Pakistan, 2019, pp. 1–5.

[35] S. Potluri and C. Diedrich, Accelerated deep neural networks for enhanced intrusion detection system, in *Proc. 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation*, Berlin, Germany, 2016, pp. 1–8.

[36] J. Kim, J. Kim, H. L. T. Thu, and H. Kim, Long short term memory recurrent neural network classifier for intrusion detection, in *Proc. 2016 International Conference on Plateform Technology and Service*, Jeju, Republic of Korea, 2016, pp. 1–5.

[37] J. Zhang, Anomaly detecting and ranking of the cloud computing platform by multi-view learning, *Multimedia Tools and Applications*, vol. 78, pp. 30923–30942, 2019.

[38] F. B. Ahmad, A. Nawaz, T. Ali, A. A. Kiani, and G. Mustafa, Securing cloud data: A machine learning based data categorization approach for cloud computing, http://doi.org/10.21203/rs.3.rs-1315357/v1, 2022.

[39] A. Mubarakali, K. Srinivasan, R. Mukhalid, S. C. Jaganathan, and N. Marina, Security challenges in Internet of things: Distributed denial of service attack detection using support vector machine-based expert systems, *Computational Intelligence*, vol. 36, no. 4, pp. 1580–1592, 2020.

[40] N. M. Abdulkareem and A. M. Abdulazeez, Machine learning classification based on radom forest algorithm: A review, *International Journal of Science and Business*, vol. 5, no. 2, pp. 128–142, 2021.

[41] L. Breiman, Random forests, *Machine Learning*, vol. 45, pp. 5–32, 2001.

[42] I. Reis, D. Baron, and S. Shahaf, Probabilistic random forest: A machine learning algorithm for noisy data sets, *The Astronomical Journal*, vol. 157, no. 1, p. 16, 2018.

[43] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, Random forests and decision trees, *IJCSI International Journal of Computer Science Issues*, vol. 9, no. 5, pp. 272–278, 2012.

[44] B. O. Yigin, O. Algin, and G. Saygili, Comparison of morphometric parameters in prediction of hydrocephalus using random forests, *Computers in Biology and Medicine*, vol. 116, p. 103547, 2020.

[45] 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.

[46] A. Devarakonda, N. Sharma, P. Saha, and S. Ramya, Network intrusion detection: A comparative study of four classifiers using the NSL-KDD and KDD'99 datasets, *Journal of Physics: Conference Series*, vol. 2161, p. 012043, 2022.

[47] M. Zeeshan, Q. Riaz, M. A. Bilal, M. K. Shahzad, H. Jabeen, S. A. Haider, and A. Rahim, Protocol-based deep intrusion detection for DoS and DDoS attacks using UNSW-NB15 and Bot-IoT data-sets, *IEEE Access*, vol.10, pp. 2269–2283, 2021.

[48] M. Shafiq, Z. Tian, A. K. Bashir, X. Du, and M. Guizani, CorrAUC: A malicious Bot-IoT traffic detection method in IoT network using machine-learning techniques, *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3242–3254, 2021.

[49] M. Shafiq, Z. Tian, Y. Sun, X. Du, and M. Guizani, Selection of effective machine learning algorithm and Bot-IoT attacks traffic identification for Internet of things in smart city, *Future Generation Computer Systems*, vol. 107, pp. 433–442, 2020.

[50] M. Hossin and M. N. Sulaiman, A review on evaluation metrics for data classification evaluations, *International Journal of Data Mining & Knowledge Management Process*, doi: 10.5121/ijdkp.2015.5201.

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, artifical intelligence, intrusion detection, and smart cities. He is also a reviewer of various scientific journals.

Hanaa Attou received the engineer diploma in operational research and decision support from National Institute of Statistics and Applied Economics, Morocco in 2020. She is currently pursuing the PhD degree of computer security at Cadi Ayyad University, Marrakech. Her main field of research interest is machine learning, intrusion detection, and cloud environment security.

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 the reviewer of various scientific journals.

Mourade Azrour received the PhD degree from Moulay Ismail University of Meknès, Errachidia, Morocco in 2019, 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, and 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*.

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, and a local publishing and research coordinator of 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.