Hindawi
Security and Communication Networks
Volume 2021, Article ID 1230593, 8 pages
https://doi.org/10.1155/2021/1230593

Research Article

A Reliable Network Intrusion Detection Approach Using Decision Tree with Enhanced Data Quality

Azidine Guezzaz,[1] **Said Benkirane**,[1] **Mourade Azrour**,[2] **and Shahzada Khurram**[3]

[1]*Computer Sciences Department, Technology Higher School, Cadi Ayyad University, Essaouira, Morocco*
[2]*Computer Sciences Department, IDMS Team, Faculty of Sciences and Technics, Moulay Ismail University, Errachidia, Morocco*
[3]*Computer Department of Information Security, Faculty of Computing, Islamia University, Bahawalpur, Pakistan*

Correspondence should be addressed to Azidine Guezzaz; a.guzzaz@gmail.com

Received 2 July 2021; Revised 26 July 2021; Accepted 13 August 2021; Published 21 August 2021

Academic Editor: Azeem Irshad

Due to the recent advancements in the Internet of things (IoT) and cloud computing technologies and growing number of devices connected to the Internet, the security and privacy issues are important to be resolved and protect the data and computer network. To provide security, a real-time monitoring of the network data and resources is needed. Intrusion detection systems have been used to monitor, detect, and alert an intrusion event in real time. Recently, the intrusion detection systems (IDS) incorporate several machine learning (ML) techniques. One of the techniques is decision tree, which can take reliable network measures and make good decisions by increasing the detection rate and accuracy. In this paper, we propose a reliable network intrusion detection approach using decision tree with enhanced data quality. Specifically, network data preprocessing and entropy decision feature selection is carried out for enhancing the data quality and relevant training; then, a decision tree classifier is built for reliable intrusion detection. Experimental study on two datasets shows that the proposed model can reach robust results. Actually, our model achieves 99.42% and 98.80% accuracy with NSL-KDD and CICIDS2017 datasets, respectively. The novel approach gives many advantages compared to the other models in term of accuracy (ACC), detection rate (DR), and false alarm rate (FAR).

1. Introduction

The computer security threats are becoming quite challenging with the growing capabilities of the adversaries, influencing the reliability of data communication and networks. The recent advancements in cloud computing and IoT technologies enabled new attack vectors for the adversary and even more prone to attacks [1–3]. The IoT applications enable the attacks not only focusing stealing the data but can also impacting human lives. For example, a hacked home utility smart heater can be used to automatically increase the temperature and indirectly impact the human beings living in the home [4, 5]. Hence, the main goal of security is to provide integrity, confidentiality, and availability by implementing various security tools and policies that can protect data and detect attacks targeting the IoT. [4, 6]. An intrusion tries to violate one of security objectives and infects systems. Hence, many tools and methods, such as IDS, are developed to secure networks and systems from intrusions [7–9]. Thereby, intrusion detection is a set of techniques implemented to detect undesirable activities by classifying data activity into normal or intrusion [6, 8]. The intrusion detection techniques detect and stop intrusions from outside or within a monitored network.

For this reason, two fundamental detection approaches can be used. The first one is called misuse detection; it is based on a known attack signature to detect intrusion. The second one is named anomaly detection or behavioral detection, based on a deviation from a normal model [1, 8, 10]. The hybrid detection approaches combine advantages of both misuse and anomaly detection and aim to increase detection rate and accuracy of IDS [9, 11, 12]. A considerable distinction is made between network IDS (NIDS) and host IDS (HIDS)

[1, 8]. Formally, an IDS can be software or hardware which detects malicious traffic, makes accurately automatic decisions, and interrupts intrusions quickly in real time with an automatic response [6, 8].

Despite their efficiency, the IDS suffers from a number of limitations, such as real-time analysis and detection, generated alarm, and data quality, that can decrease detection rate and accuracy performances [6, 8]. Therefore, intrusion detection is still an effective and dynamic research field.

Recently, ML methods have been integrated to enhance intrusion detection and reinforce computer security. Numerous research contributions explore how to incorporate ML techniques in intrusion detection to obtain reliable IDS with accurate performances by enhancing data quality and training [13–20]. The decision tree is an induction algorithm which has been used for classification in many issues. It is based on splitting features and testing the value of each one. The splitting process continues until each branch can be labelled with just one classification [21, 22]. The decision tree is more than equivalent representation to the training set. Hence, it can be used to predict the values of other instances not in the training set. The decision tree is widely used as a mean of generating classification rules because of the existence of a simple but very powerful algorithm called Top-Down Induction of Decision Trees (TDIDT). It is guaranteed to give a decision tree that correctly corresponds to the data provided by two of the best known being ID3 and C4.5 [22].

On the other side, the data is not always obtained in a structured form. For relevant analysis, the unstructured data have to be preprocessed. This operation is an essential stage which performed to enhance data quality and make accurate decisions. Data quality techniques are implemented before training and classification process [17, 23, 24]. Besides, feature selection is a desirable process aiming to select the useful features to both reduce the computational cost of modelling and to improve the performance of the predictive model [13, 24].

In this paper, we propose a novel network intrusion detection approach based on the decision tree method to train and build a binary classifier model and make accurate decisions. The features' engineering techniques were used to improve the data quality. Experimental results on the NSL-KDD dataset and CICIDS2017 dataset demonstrate that our proposed approach gives good performances in terms of accuracy DR and FAR. Two main contributions have been validated in this research work. Firstly, we implement feature selection using entropy decision technique to improve data quality. Secondly, we build a classifier model based on decision tree algorithm to achieve effective network intrusion detection approach.

The remainder of this paper is organized as follows. Section 2 presents related work on intrusion detection, especially which integrated ML techniques to improve IDS performances. Section 3 describes in detail the proposed solutions for the novel approach. In Section 4, we discuss experimental results, performance of the proposed model, and its comparison with other models. Finally, the conclusion and future works presented in Section 4.

2. Related Works

During the last decade, a set of contributions of intrusion detection were adopted in [8, 10, 11, 17, 21, 25, 26] to ensure computer security objectives. The research in intrusion detection is oriented towards on automatic response to increase effectiveness and capability of IDS [6]. Therefore, to obtain reliable IDS, the false positive (FP) and false negative (FN) rate should be low, but also, true positive (TP) and true negative (TN) rate should be high. Furthermore, including ML techniques in intrusion detection becomes an excited research domain [13–20]. Hence, intrusion detection based on ML is a classification task aiming to detect intrusions using labelled data by building a classifier able to distinguish between normal and abnormal activity [11, 16, 21, 27, 28]. Several ML techniques, such as decision tree [21], random forest [29], nearest neighbour [30], Naïve Bayes [26, 27], support vector machine [17], fuzzy clustering [15], reinforcement based learning [19], and deep learning methods [1, 6, 14, 18, 25, 26, 31, 32] have been integrated to enhance IDS by discovering knowledge from intrusion detection datasets [9, 31, 33, 34]. For more improvements, a set of feature engineering techniques, such as feature selection, are made to enhance data quality. They allow a relevant data process used to train and build effective classifier [13, 17, 23, 25, 35, 36].

In 2018, Karami [37] proposed an anomaly-based intrusion detection system using the fuzzy SOM method. In 2020, Tabash et al. [26] proposed an intrusion detection model which integrated NB and DL technique. The model implemented genetic algorithm for a good feature selection. In 2015, Ghazali et al. [27] proposed a detection model for intrusive communication. This research work tests five classification techniques: SimpleCart, NB, BFTree, PART, and Ridor. The performances' measures on NSL-KDD dataset demonstrate ACC 96.7%, DR 95.5%, and FAR 4.7%. In 2017, Kevric et al. [28] proposed a combining classifier approach using tree algorithm for network intrusion detection. The model is evaluated on NSL-KDD dataset ACC 89.24%. In 2018, Hadi [29] proposed a model based on random forest algorithm for selecting a significant feature. The model was evaluated using NSL-KDD. The results of the proposed model are ACC 99.33%, DR 0.993% TP, and FAR 0.001% FP. In 2019, Gu et al. [17] proposed a model of an ensemble SVM-based intrusion detection with LMDRT transformation as an effective method to enhance data quality. The performances' results on CICIDS2017 dataset are ACC 93.64%, DR 97.56%, and FAR 20.28%. In 2020, Elmasry et al. [32] developed a DL model for network intrusion detection using a double PSO metaheuristic. The model is evaluated on CICIDS2017 dataset and gives ACC 92.92%, DR 92.38%, and FAR 3.24%. In 2019, Prasard et al. [36] proposed new IDS which works on subset of features by extracting significant features using the probabilistic method. The BRS method is implemented to categorize samples into normal, intermediary, and abnormal category based on the rough set. The model is trained and tested on CICIDS2017 dataset and demonstrates ACC 97.6%, DR 96.38%, and FAR 3.00%. In 2019, Ahmim et al. [21] proposed a hybrid IDS model which combines the classifier model based on decision tree, REP tree, JRIP algorithm, and forest PA. The performances of the novel model are evaluated using CICIDS2017 dataset and presented ACC 96.66%, DR 94.475%, and FAR 4.47%.

From the state-of-the-art literature survey, it is proven that the learning methods and data quality are two useful tasks which determine the robustness of IDS [6, 17, 26–29, 32, 36, 37]. These research works implement much of techniques for a high quality of data by not only reducing and selecting features but also building improved classifiers to better categorize data activities.

3. Novel Network Intrusion Detection Approach

In this section, we describe our methodology and proposed solutions aiming to implement and validate the novel approach. By enhancing feature engineering and classification techniques, we obtained reliable IDS with accurate performances.

3.1. Our Proposed Model. As depicted in Figure 1, the proposed model consists of three main components including data quality component, building of classifier component, and intrusion detection deployment component. The details of those three components are given in the following.

Part 1: data quality process.

The main goal of this component is collecting and preprocessing the data. Hence, the system executes the process that can gather and accumulate necessary data from networks. Once the data are collected, a specific data preprocessing is performed on gathered network traffic. The data preprocessing portion evaluates the data and ignores the incompatible data types. Furthermore, the data is sanitized and the resulting data is saved. In addition, the data is transformed and the features of network dataset are finalized. We used the entropy decision technique to select the features.

Part 2: building of the classifier.

Once the first part is completed, the second one is started. Generally, the objective of second part, as it is clear in its name, is to build a classifier model. The input here is the transformed data obtained in the data quality process part. In the classifier building part, we can distinguish between two main phases: model training phase and model validation phase. In the first phase, three portions of data are used for training a decision tree classifier implemented in our proposed approach. Then, in the second phase, the rest of data are used to validate our model.

Part 3: network intrusion detection deployment.

After building of the classifier model, the third part comes for deploying the network intrusion detection. At this point, actual tests are necessary to improve the performance of reliable IDS. Hence, we are in aptitude to check its capacity to classify activities in normal or abnormal. So, based on the classification results, the IDS can made accurate.

3.2. Description of Proposed Solutions. As we mentioned above, the first step which is made by our approach is to collect and transform data with feature selection according to needs of analysis and detection. The data quality is an important and essential task to train and build an accurate intrusion detection model. Hence, this step aims to prepare data for analysis and make accurate decision. We start first with data transformation by applying feature selection using entropy decision on original traffic collected within network traffic to obtain a good training set. In fact, it is a critical step aiming to improve accuracy of our approach. It aims also to overcome training complexity by reducing analysed data and obtain a great model with best performances in terms of accuracy, detection rate, and real-time detection. A particular preprocessing is applied on collected network traffic before the analysis step. Data normalization is performed. For this, we suggest and implement a particular coding to enumerate feature values and establish a pattern of activities facilitating the distinction between the activities. The goal of the feature extraction is to reduce the number of features in collected data from networks. It aims to summarize most of the information contained in this original data by creating new features. The feature selection aims instead to choose the important existing features in the original data and discard less important ones. For this reason, we use entropy decision technique for feature selection. The implementation of components that constitute our approach is described in Figure 2.

We obtain a transformed data by implementing proposed data quality techniques, aiming to increase our approach accuracy. This allows training and validating of an effective intrusion detection model based on the decision tree to make relevant decisions in real time. Moreover, intrusion detection is considered as a classification task aiming to classify incoming traffic in normal activity or intrusion. Hence, the main objective of this part is to predict a binary value to validate the classifier able to answer question with a yes or a no. Thus, we encoded both classes in numerical variable: +1 for normal activity and −1 for intrusion. We remember that the number of features must be fixed in advance. For the validation step of our model, there are various strategies used to split the data into a training and test set. In this case, we use the efficient and recommended one, k-fold [1].

According to standard components of an IDS mentioned in [8, 29], our approach is constituted by four parts: data collection part, preprocessing part, decision-making part, and response part. The proposed approach focuses on the preprocessing part by improving data quality technique used to train and build an accurate classifier which is able to discover intrusions within traffic network. It focuses also on enhancing the decision-making part by integrating the decision-tree classifier. A set of research works have been made in [6, 13, 24] to improve others parts of IDS, such as

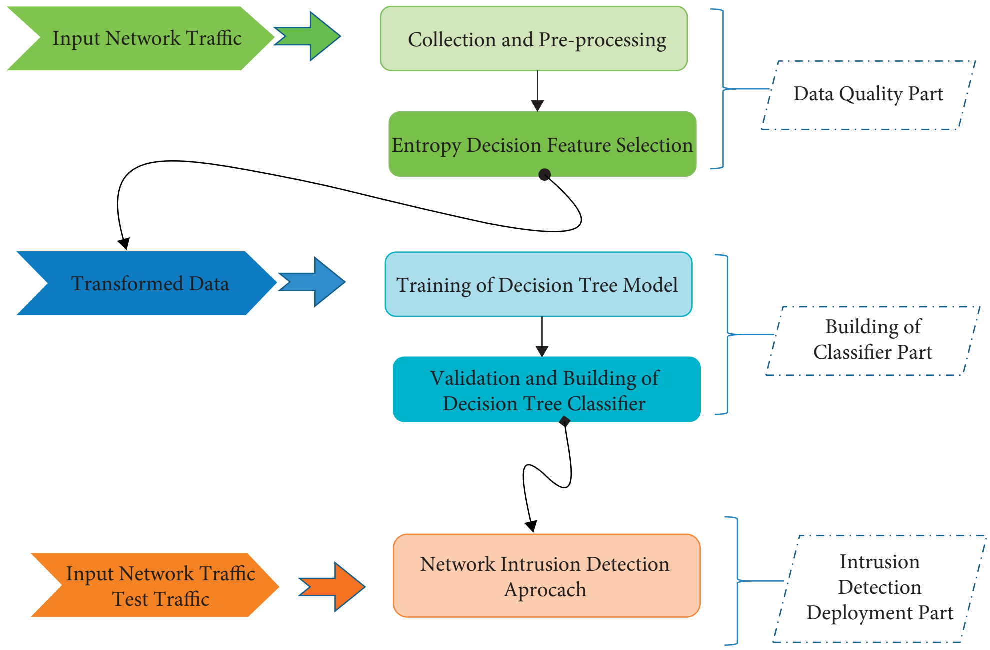


FIGURE 1: Proposed network intrusion detection model.

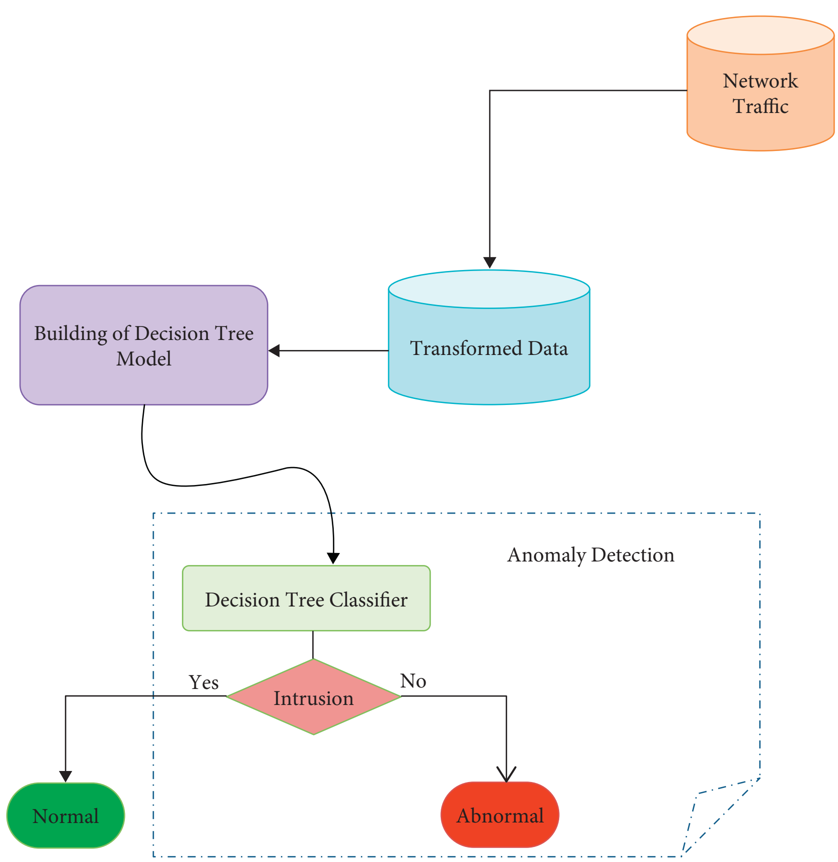


FIGURE 2: Procedure of validation and building classifier.

data collection, dimensionality reducing, and real-time response which are not taken into account in this research work.

4. Experimental Results and Discussion

4.1. Dataset Description. The assessment of datasets plays a vital role in validation of intrusion detection approaches. Therefore, for evaluating any IDS using ML techniques, one can select the desired dataset among a large number of appropriate and available datasets. For instance, numerous public datasets are available [9, 31, 33, 34] and can be used freely for evaluation proposed methods' capability. In our case, we have selected two types of datasets including NSL-KDD and CICIDS2017, which are used for training and performances' evaluation and validation of the proposed approach.

The NSL-KDD dataset was created from KDD cup 99 dataset [9, 27]. It contains 125,973 records of the training set and 22,544 for the test set. It has 22 training instances' attacks and 41 features in which 21 of them describe connection itself and 19 for nature of connection of the same host [33, 38]. The novelty and instances' volume of the NSL-KDD dataset make it very practical. On the contrary, the CICIDS2017 dataset was created from Canadian Institute for Cyber Security. It aims to overcome the limitations of the actual dataset and present an effective dataset for intrusion detection. It is a labelled dataset that comprises behavior and new malware attacks and is consisted of 8 files containing 2,830,743 instances. The CICIDS2017 dataset integrates 80 features' network flow captured at July 2017 from network traffic using CICFlowMeter tool [9].

Those two used datasets in this research work, NSL-KDD dataset and CICIDS2017 dataset, are available at [39, 40], respectively.

4.2. Experiments' Environment. The experimental setting of our research work is performed and evaluated on a computer with a Core-i7 2700K CPU@ 2.50 GHz and 32 GB of DDR3 running windows 7 professional 64 bits. The entropy feature selection and decision-tree model training are implemented using python version 3.8.0.

To validate our proposed intrusion detection model, we use the 10-fold cross-validation technique to obtain the training and test set. Hence, we split randomly full dataset into ten parts with the same size. Nine parts are used in the training and the last part in the test step. Finally, the performances of the model are presented by repeating this procedure ten times.

4.3. Data Transformation. In the implementation step, we propose to extract samples of dataset to avoid some drawbacks such as processing and big volume of data. The data extraction from each used dataset is given in Table 1.

Feature selection is a relevant technique included by our network intrusion detection approach. It is implemented and incorporated to select useful features for reliable detection and decision-making. For this, we implement entropy decision technique.

Table 1: Data extraction from NSL-KDD and CICIDS2017 datasets.

	Category	Original size	Extracted size
NSL-KDD dataset	Training	125,973	25,195
	Test	22,544	4,509
	Total	148,517	29,704
CICIDS2017 dataset	Benign	2,273,097	113,655
	Attack	557,646	27,883
	Total	2,830,743	141,538

The encoding step is performed to assign numeric values to categorical features for making relevant processing. To avoid undesirable influence problem of high weights, we normalize continuous features values. Equation (1) is used to find the new value. Hence, we make the values of each feature run from 0 to 1. If the lowest value of a given feature x is min and the highest value is max, we convert each value of x to

$$\frac{(\text{value}\,(x) - \text{min})}{(\text{max} - \text{min})}. \tag{1}$$

Furthermore, all continuous features are in range [0, 1].

4.4. Metrics Evaluation and Discussion. The most obvious criterion to use for estimating the performances of a classifier is predictive accuracy. The proportion of a set of unseen instances that it correctly classifies. For numerical performances' evaluation of the proposed model, the following metrics are used.

These metric performances are not dependent on the size of the training and test set and can be really helpful in assessing the performance of the full model. Based on the confusion matrix (Table 2), the performances' metrics are calculated.

ACC is obtained from equation (2). It is the ratio of instances that are correctly predicted as normal or attack to the overall number of instances in the test set:

$$ACC = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{2}$$

DR is calculated using equation (3) and indicates the ratio of the number of instances that are correctly classified as attack to the total number of attack instances present in the test set:

$$DR = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{3}$$

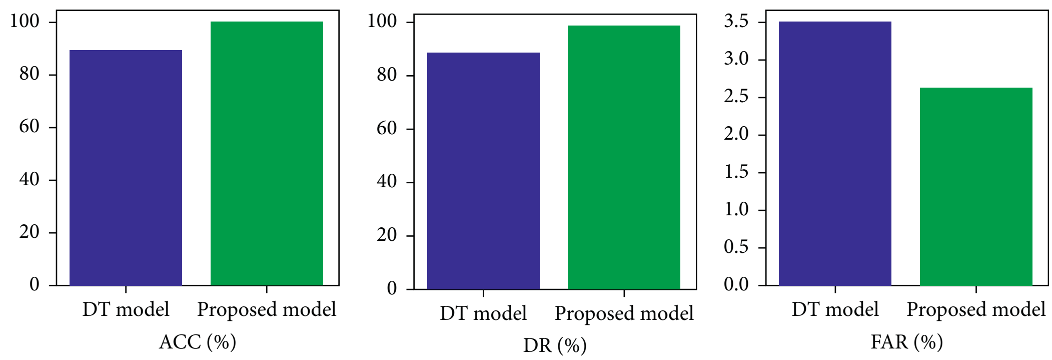
FAR is obtained from equation (4) and represents the ratio of instances which is categorized as attack to the overall number of instances of normal behavior:

$$FAR = \frac{\text{FP}}{\text{FP} + \text{TN}}. \tag{4}$$

In this research work, we start with comparing detection assessment of our proposed model for novel approach and decision-tree model only. The results shown in Figures 3 and 4 demonstrate this comparison according to ACC, DR, and FAR on the NSL-KDD dataset and the CICIDS2017 dataset.

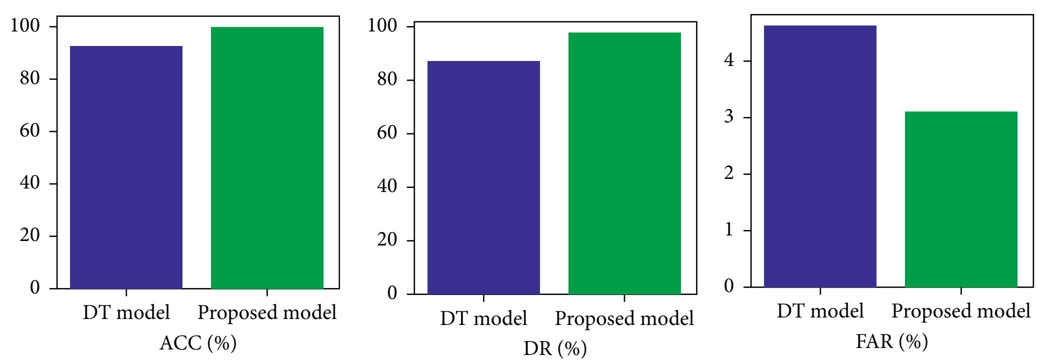
Table 2: Confusion matrix.

Actual class	Predicted class	
	Attack	Normal
Attack	TP	FN
Normal	FP	TN



(a) (b) (c)

Figure 3: (a) ACC results of the DT model and our proposed model on the NSL-KDD dataset. (b) DR results. (c) FAR results.



(a) (b) (c)

Figure 4: (a) ACC results of the DT model and our proposed model on the CICIDS2017 dataset. (b) DR results. (c) FAR results.

Figures 3(a) and 4(a) show that accuracy of the proposed model is specifically better than the model based on the decision tree only. Figures 3(b) and 4(b) demonstrate the DR of both IDS. It validates that the DR of the proposed IDS model is higher than the IDS based on the decision tree only on the NSL-KDD dataset and the CICIDS2017 dataset.

The results demonstrated above are summarized in Tables 3 and 4. They show that our proposed model can reach significant performances than the decision tree only. For the NSL-KDD dataset, the ACC of our proposed model achieves 99.42%, while the decision tree only exceeds 89%. In terms of DR and FAR, our proposed model obtains 98.2% and 2.64%, respectively, while the decision tree only presents DR 88.5% and FAR 3.5%. For the CICIDS2017 dataset, our proposed model indicates high performances in terms of ACC 98.8%, DR 97.3%, and FAR 3.10%. Besides, the decision tree only gives ACC 92%, DR 86.7%, and FAR 4.6%.

Table 3: Performances' metrics of the decision tree and the proposed model using the NSL-KDD dataset.

	ACC (%)	DR (%)	FAR (%)
Decision tree	89.00	88.50	3.50
Proposed approach	99.42	98.20	2.64

Table 4: Performances' metrics of the decision tree and the proposed model using the CICIDS2017 dataset.

	ACC (%)	DR (%)	FAR (%)
Decision tree	92.00	86.70	4.60
Proposed approach	98.80	97.30	3.10

The results obtained validate that our approach gives great detection capability in terms of ACC, DR, and FAR. Specifically, they demonstrate that the performances'

Table 5: Performances' comparison with other models on NSL-KDD.

	Method	Accuracy (%)	DR (%)	FAR (%)
Masdari and Khezri [12]	Five classification	96.70	95.50	4.70
Ahmim et al. [21]	Tree algorithm	89.24	—	—
Fang [16]	RF	99.33	0.993 TP	0.001FP
Proposed approach	DTE	99.42	98.20	2.64

Table 6: Performances' comparison with other models on CICIDS2017.

	Method	Accuracy (%)	DR (%)	FAR (%)
Chiba et al. [1]	DTRM	96.66	94.475	4.47
Alazzam et al. [13]	EnSVM	93.64	97.56	20.28
Ayo et al. [25]	BRS	97.96	96.38	3.00
Khraisat et al. [9]	DL	92.92	92.38	3.24
Proposed approach	DTE	98.80	97.30	3.10

metrics of our proposed model are higher *o* NSL-KDD dataset but low CICIDS2017 dataset. According to the evaluation performances, our proposed IDS model can reach great performances. The comparison with the model which uses the decision tree only indicates the effectiveness of our network intrusion detection approach.

Concretely, our proposed intrusion detection model is specified by high performances of ACC, DR, and FAR. Furthermore, we perform a comparison between our IDS and other recent intrusion detection approaches based on the NSL-KDD dataset and the CICIDS2017 dataset. Typically, the recent works that integrate ML techniques are tree algorithm, RF, DTRM, EnSVM, BRS, and DL. The comparison results are presented in Tables 5 and 6.

From the obtained results, we conclude that our proposed IDS approach is relevant, achieves important performances, and gives relevant training by implementing fast data quality techniques. Using the NSL-KDD dataset and the CICIDS2017 dataset, it is proven that our approach is reliable and reaches good results compared with other models. The novel approach can be integrated and used to secure various environments such as IoT environment and cloud computing.

5. Conclusion and Future Works

Intrusion detection is a set of enhanced techniques implemented to monitor systems and data to be more secure. In this paper, we present a reliable network intrusion detection approach based on decision-tree classifier and engineering feature techniques. According to heterogeneity of data, a preprocessing phase is setting up to increase detection rate and accuracy of IDS. Also, a feature selection technique based on the entropy decision-tree method is handled before building the model for high data quality. The validation of novel approach is achieved by proposed solutions that guarantee an efficient accuracy. The performances are evaluated on two datasets: NSL-KDD and CICIDS2017. Hence, the novel proposed network intrusion detection approach presents many advantages and provides high accuracy compared with other models. The future works will integrate other efficient ML techniques such as deep learning in various parts to empower detection rate and accuracy of our approach.

Data Availability

The assessments and experimental results, obtained using Anaconda 3 IDE, are available at https://sites.google.com/umi.ac.ma/azrour.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Z. Chiba, N. Abghour, K. Moussaid, A. El omri, and M. Rida, "Intelligent approach to build a deep neural network based IDS for cloud environment using combination of machine learning algorithms," *Computers & Security*, vol. 86, pp. 291–317, 2019.

[2] A. Irshad, S. A. Chaudhry, O. A. Alomari, K. Yahya, and N. Kumar, "A novel pairing-free lightweight authentication protocol for mobile cloud computing framework," *IEEE Systems Journal*, vol. 2020, Article ID 2998721, 2020.

[3] S. A. Chaudhry, I. L. Kim, S. Rho, M. S. Farash, and T. Shon, "An improved anonymous authentication scheme for distributed mobile cloud computing services," *Cluster Computing*, vol. 22, no. S1, pp. 1595–1609, 2019.

[4] A. Irshad, M. Usman, S. A. Chaudhry, H. Naqvi, and M. Shafiq, "A provably secure and efficient authenticated key agreement scheme for energy internet-based vehicle-to-grid technology framework," *IEEE Transactions on Industry Applications*, vol. 56, no. 4, pp. 4425–4435, 2020.

[5] S. A. Chaudhry, M. S. Farash, N. Kumar, and M. H. Alsharif, "PFLUA-DIoT: a pairing free lightweight and unlinkable user access control scheme for distributed IoT environments," *IEEE Systems Journal*, vol. 2020, Article ID 3036425, 2020.

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

[7] G. Fernandes, J. J. P. C. Rodrigues, and L. F. Carvalho, "A comprehensive survey on network anomaly detection," *Telecommunication Systems*, vol. 70, pp. 447–489, 2019.

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

[9] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges," *Cybersecurity*, vol. 2, 2019.

[10] S.-Y. Ji, B.-K. Jeong, S. Choi, and D. H. Jeong, "A multi-level intrusion detection method for abnormal network behaviors,"

Journal of Network and Computer Applications, vol. 62, pp. 9–17, 2016.

[11] Ü. Çavuşoğlu, "A new hybrid approach for intrusion detection using machine learning methods," *Applied Intelligence*, vol. 49, pp. 2735–2761, 2019.

[12] M. Masdari and H. Khezri, "A survey and taxonomy of the fuzzy signature-based intrusion detection systems," *Applied Soft Computing*, vol. 92, Article ID 106301, 2020.

[13] H. Alazzam, A. Sharieh, and K. E. Sabri, "A feature selection algorithm for intrusion detection system based on Pigeon Inspired Optimizer," *Expert Systems with Applications*, vol. 148, Article ID 113249, 2020.

[14] A. Aldweesh, A. Derhab, and Z. E. Ahmed, "Deep learning approaches for anomaly-based intrusion detection systems: a survey, taxonomy, and open issues," *Knowledge-Based Systems*, vol. 189, Article ID 105124, 2020.

[15] M. Amini, J. Rezaeenour, and E. Hadavandi, "A neural network ensemble classifier for effective intrusion detection using fuzzy clustering and radial basis function networks," *The International Journal on Artificial Intelligence Tools*, vol. 25, no. 2, 2016.

[16] W. Fang, X. Tan, and D. Wilbur, "Application of intrusion detection technology in network safety based on machine learning," *Safety Science*, vol. 124, Article ID 104604, 2020.

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

[18] M. M. Hassan, A. Gumaei, A. Alsanad, M. Alrubaian, and G. Fortino, "A hybrid deep learning model for efficient intrusion detection in big data environment," *Information Sciences*, vol. 513, pp. 386–396, 2020.

[19] K. Sethi, E. Sai Rupesh, R. Kumar, P. Bera, and Y. Venu Madhav, "A context-aware robust intrusion detection system: a reinforcement learning-based approach," *International Journal of Information Security*, vol. 19, no. 6, pp. 657–678, 2020.

[20] A. Sommer and V. Paxson, "Outside the closed world: on using machine learning for network intrusion detection,," in *Proceedings of the 2010 IEEE Symposium on Security and Privacy*, pp. 305–316, Oakland, May 2010.

[21] A. Ahmim, L. Maglaras, M. A. Ferrag, M. Derdour, and H. Janicke, "A novel hierarchical intrusion detection system based on decision tree and rules-based models,"pp. 228–233, https://ieeexplore.ieee.org/xpl/conhome/8790388/proceeding, Santorini, Greece, May 2019.

[22] C. Guo, Y. Ping, N. Liu, and S.-S. Luo, "A two-level hybrid approach for intrusion detection," *Neurocomputing*, vol. 214, pp. 391–400, 2016.

[23] K. Jeyakumar, T. Revathi, and S. Karpagam, "Intrusion detection using artificial neural networks with best set of features," *The International Arab Journal of Information Technology*, vol. 12, no. 6A, 2015.

[24] M. Rostami, K. Berahmand, E. Nasiri, and S. Forouzandeh, "Review of swarm intelligence-based feature selection methods," *Engineering Applications of Artificial Intelligence*, vol. 100, Article ID 104210, 2021.

[25] F. E. Ayo, S. O. Folorunso, A. A. Abayomi-Alli, A. O. Adekunle, J. B. Awotunde, and J. B. Awotunde, "Network intrusion detection based on deep learning model optimized with rule-based hybrid feature selection," *Information Security Journal: A Global Perspective*, vol. 29, no. 6, pp. 267–283, 2020.

[26] M. Tabash, M. Abd Allah, and B. Tawfik, "Intrusion detection model using naive bayes and deep learning technique," *The International Arab Journal of Information Technology*, vol. 17, no. 2, 2020.

[27] A. Ghazali, W. Nuaimy, A. Al-Atabi, and I. Jamaludin, "Comparison of classification models for Nsl-Kdd dataset for network anomaly detection," *Academic Journal of Science*, vol. 4, no. 1, pp. 199–206, 2015.

[28] J. Kevric, S. Jukic, and A. Subasi, "An effective combining classifier approach using tree algorithms for network intrusion detection," *Neural Computing & Applications*, vol. 28, no. S1, pp. 1051–1058, 2017.

[29] A. Hadi, "Performance analysis of big data intrusion detection system over random forest algorithm," *International Journal of Applied Engineering Research*, vol. 13, no. 2, pp. 1520–1527, 2018.

[30] A. Topîrceanu and G. Grosseck, "Decision tree learning used for the classification of student archetypes in online courses," *Procedia Computer Science* in *Proceedings of the 21st International Conference on Knowledge Based and Intelligent Information and Engineering*, vol. 112, pp. 51–60, Marseille, France, September 2017.

[31] 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, Article ID 102419, 2020.

[32] W. Elmasry, A. Akbulut, and A. H. Zaim, "Evolving deep learning architectures for network intrusion detection using a double PSO metaheuristic," *Computer Networks*, vol. 168, Article ID 107042, 2020.

[33] N. Moustafa and J. Slay, "The evaluation of network anomaly detection systems: statistical analysis of the unsw-nb15 data set and the comparison with the kdd99 data set," *Information Security Journal: A Global Perspective*, vol. 25, no. 1-3, pp. 18–31, 2016.

[34] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *Proceedings of the 4th International Conference on Information Systems Security and Privacy*, pp. 108–116, Madeira, Portugal, January 2018.

[35] M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, "Building an intrusion detection system using a filter-based feature selection algorithm," *IEEE Transactions on Computers*, vol. 65, no. 10, pp. 2986–2998, 2016.

[36] M. Prasad, S. Tripathi, and K. Dahal, "An efficient feature selection based Bayesian and rough set approach for intrusion detection," *Applied Soft Computing*, vol. 2020, Article ID 105980, 2020.

[37] A. Karami, "An anomaly-based intrusion detection system in presence of benign outliers with visualization capabilities," *Expert Systems with Applications*, vol. 108, pp. 36–60, 2018.

[38] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 dataset," in *Proceedings of the Second 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, pp. 1–6, Ottawa, Canada, July 2009.

[39] https://web.archive.org/web/20150205070216/http://nsl.cs.unb.ca/NSL-KDD/ (2021, July 24).

[40] https://www.unb.ca/cic/datasets/ids-2017.html (2021, July 24).