Ingénierie des Systèmes d'Information
Vol. 28, No. 4, August, 2023, pp. 1003-1012
Journal homepage: http://iieta.org/journals/isi

Employing Hybrid ANOVA-RFE with Machine and Deep Learning Models for Enhanced IoT and IIoT Attack Detection and Classification

Moumena Salah Yassen[1], Raghdah Adnan Abdulrazzq[2], Ahmed Burhan Mohammed[3*]

[1] Software Department, College of Computer Science and Information Technology, Kirkuk University, Kirkuk 36001, Iraq
[2] Business Administration Department, College Administration and Economics, University of Kirkuk Al-Wasti Area Kirkuk, Kirkuk 36001, Iraq
[3] College of Dentistry, University of Kirkuk, Kirkuk 36001, Iraq

Corresponding Author Email: ahmedlogic79@uokirkuk.edu.iq

https://doi.org/10.18280/isi.280420

Received: 18 May 2023
Revised: 26 July 2023
Accepted: 11 August 2023
Available online: 31 August 2023

Keywords:
attack detection, Internet of Things (IoT) security, feature selection, machine learning, deep learning

ABSTRACT

The Internet of Things (IoT) has become an integral component in various applications, with significant prominence in healthcare and cybersecurity sectors. It is indispensable in medical diagnostics, monitoring, decision-support systems, and the safeguarding of sensitive data. However, the traditional methodologies have shown limitations in their ability to detect and classify all types of attacks effectively. This study presents a robust feature selection model, ANOVA-Recursive Feature Elimination (ANOVA-RFE), implemented with both Machine and Deep Learning paradigms, aiming to augment the security level by enhancing attack detection and classification. The models were trained using both the entire feature set and the selected features identified by ANOVA-RFE, demonstrating the efficiency and precision of the proposed method. The experiments yielded an accuracy of 100% and 99.96% using only the top five selected features from the first and second datasets, respectively. Furthermore, the performance of Gaussian Naive Bayes (GNB), K-Nearest Neighbors (K-NN), Random Forest (RF), AdaBoost (AB), Logistic Regression (LR), Decision Tree (DT), and Long Short-Term Memory (LSTM) models are evaluated, showcasing their respective accuracies on the first dataset. A score-level fusion was also employed, and the results were benchmarked against the current state-of-the-art, validating the robustness and high precision of the current study. Future work should consider analyzing different datasets and addressing further challenges.

1. INTRODUCTION

The proliferation of the Internet of Things (IoT) technology has permeated various facets of modern life [1]. Concomitant with this growth, cybersecurity threats have escalated across diverse technology-driven sectors, encompassing healthcare applications, smartphones, industrial institutions, and security platforms. This surge in threats necessitates robust artificial intelligence (AI) solutions for their prevention and detection. The traditional AI methodologies, however, have demonstrated shortcomings in coping with the swiftly expanding volume of monitoring data. Conventional rule-based systems and machine learning algorithms such as K-Nearest Neighbor (K-NN), Decision Tree (DT), and Logistic Regression (LR) are adversely affected by high dimensionality, necessitating effective feature selection to minimize training time.

IoT companies manufacture a multitude of large-scale IoT devices considering the distinct characteristics of cloud computing, particularly real-time processing. A myriad of technologies, including healthcare IoT, big data analysis, smart technology, and Industrial IoT (IIoT), are integrated into cybersecurity systems [2]. These environments call for rigorous monitoring to ensure the security of crucial information exchanged within these systems. To address this exigency, numerous intrusion detection studies have been introduced. IoT-based security systems necessitate specialized processes to protect information and impede attackers from compromising critical data [3]. Recent advancements in AI technologies have unveiled improved encryption approaches that safeguard privacy, detect intrusions, and classify threat types.

Several Machine Learning (ML) and Deep Learning (DL) algorithms have been introduced in the fields of intrusion detection, threat detection, malware detection, and ransomware detection in IoT applications [4]. Some of these rely on traditional methods, while others enhance security levels using more accurate and powerful AI technologies. The remainder of this paper is structured as follows: the following section reviews recent related work, then the aim and objectives of the study are defined. The materials and proposed methodology are subsequently detailed. Extensive experiments are explicated in the results and discussion sections, and the conclusion summarizes the overall work and proposes future directions.

2. LITERATURE REVIEW AND PROBLEM STATEMENT

The literature on botnet attack detection and mitigation provides several noteworthy contributions. Vinayakumar et al.

[5] developed a framework that utilized a two-tier environment for monitoring Domain Names and DNS log files. Their approach, which leveraged a deep learning architecture, yielded a notable decrease in the rate of false alarms and achieved an accuracy of 99.2% and 89.9% on DS1-V1 and DS2-V2 datasets, respectively. Similarly, Popoola et al. [6] proposed a system based on the LAE-BiLSTM architecture for botnet attack detection and achieved an accuracy of 91.89% on the BotIoT dataset.

In the context of IoT within healthcare, Hussain et al. [7] introduced a framework for malware traffic detection. Their methodology involved the use of IoT-Flock software to gather information about normal and malicious IoT devices. The Random Forest (RF) algorithm emerged as the most effective among the Machine Learning (ML) algorithms tested, achieving an accuracy of 99.51%.

Ferrag et al. [8] contributed to the field by introducing a new cyber security dataset, named "Edge-IIoTset Cyber Security Dataset of IoT & IIoT," derived from IoT and IIoT applications. The study produced a high-dimensional dataset of 1176 features, which was subsequently reduced to 61 features. Among the ML classifiers used to develop the attack detection system, the RF attained 80.83% accuracy, and the Support Vector Machine (SVM) reached an accuracy of 77.61%. The Deep Neural Network (DNN) architecture achieved 94.67% accuracy in a multi-class classification scenario involving 15 classes and 96.01% accuracy in a 6-class scenario.

Federated Learning-based intrusion detection systems have also been examined. Tang et al. [9] reported higher accuracy results with the Federated Learning-based method compared to traditional ones when tested on the CICID2017 intrusion dataset.

Further significant contributions have been made in the field. Bahadoripour et al. [10] applied a multi-modal deep learning model to detect attacks in cybersecurity networks. The model exhibited superior performance compared to its predecessors, recording a precision of 99%, a recall of 98%, and an F1-score of 98% on a dataset derived from the Secure Water Treatment system.

In another noteworthy study, Hussein et al. [11] proposed a system that employs fuzzy logic to compute the anomaly score of each data point. This system leverages various outlier factor methods, namely, the local outlier factor (LOF), the connectivity-based outlier factor (COF), and the generalized LOF. This approach effectively resolved the ambiguity involved in classifying data points as outliers or inliers.

This approach was further investigated by Rashid et al. [12] in the context of industrial IoT systems, with their experiments on the Edge-IIoTset dataset yielding an accuracy of 93.92%.

Hybrid deep learning models have also been applied in intrusion detection. Hnamte and Hussain [13] combined BiLSTM and CNN architectures to develop an efficient system. The model was tested on the CICIDS2018 and Edge_IIoT datasets, achieving an accuracy of 100% and 99.64%, respectively.

Investigations into IoT network anomaly mitigation systems have yielded promising results. Alzahrani and Alzahrani [14] proposed a statistical method combining the K-NN, cumulative sum, and exponentially weighted average algorithms to detect DDoS attacks. Their evaluation of the model on the Bot-IoT dataset achieved an accuracy of 99%.

While these studies offer significant advances in attack detection, they often overlook certain aspects. A common shortcoming is the lack of emphasis on attack classification. Additionally, the issue of dimensionality reduction is frequently overlooked, leading potentially to unreliable and high-cost systems. Moreover, the generalizability of these methodologies is often limited due to their validation on a single type of IoT application.

This study aims to address these gaps. It employs two different types of datasets (healthcare and cybersecurity), both of which consist of high-dimensional big data. Additionally, a novel feature selection method is proposed to reduce the complexity of the features, enhance system performance, and minimize the cost.

3. THE AIM AND OBJECTIVES Of THE STUDY

This study aims to build an attack detection and classification system in IoT healthcare and cyber security applications using feature selection, machine learning and deep learning models. The main objectives can be concluded as follows:

1. To investigate the feature selection algorithms and their effect on attack detection and classification performance in IoT applications and show the effect of dimensionality reduction on the performance of ML and DL models in attack detection problem.
2. To build and evaluate a hybrid ANOVA-RFE feature selection algorithm to combine the advantages of feature-based and wrapper-based feature selection algorithms and minimize the ML and DL models' training time with preserving a high accuracy.
3. To compare the performance of ANOVA-RFE using two high-dimensional big datasets (with different specifications in order to validate the proposed method under different challenges).
4. To evaluate the proposed feature selection methodology using different performance metrics, including computational time, accuracy, precision, recall and F1-score.

4. MATERIALS AND METHODS

4.1 Datasets

In the current study, two different datasets were proposed. Using two datasets aims to test the proposed feature selection algorithm in two different environments. The first choice of datasets includes the problem of binary classification (attack or Normal) on a high-dimensionality dataset consisting of 52 columns (51 predictors and one target) in the field of IoT-based healthcare applications [7, 15]. This dataset has a moderate size (188694 records). While the second dataset (Edge-IIoTset Cyber Security Dataset of IoT & IIoT [8]) has a more complex level since it has 63 columns, 2219200 records (big data), and 15 different categories in the target column (these categories represent the different types of attacks). This dataset contains information on the cyber security of IoT and industrial IoT applications. The normal case constitutes 58% of the records of the first dataset, while it represents 73% of the second dataset's records.

For the first dataset, the problem is a binary classification one in which there are only two cases of targets (Attack: 1, Normal: 0). While for the second dataset, the problem is a multi-class classification consisting of 15 different attacks,

including besides the normal case: DDoS UDP, DDoS HTTP, DDoS TCP, DDoS ICMP, SQL injection, password, vulnerability scanner, uploading, backdoor, fingerprinting, ransomware, XSS, MITM and port scanning attacks.

By using these two datasets, the high-dimensionality size, large records number, and the binary-vs-multiclass problems will also be discussed and compared. So, the current methodology will be evaluated under different challenging datasets to check its universality.

4.2 Methodology

Figure 1 illustrates the general steps of the proposed system applied to the first and second datasets. The main designed algorithm (hybrid feature selection method is applied to both datasets).

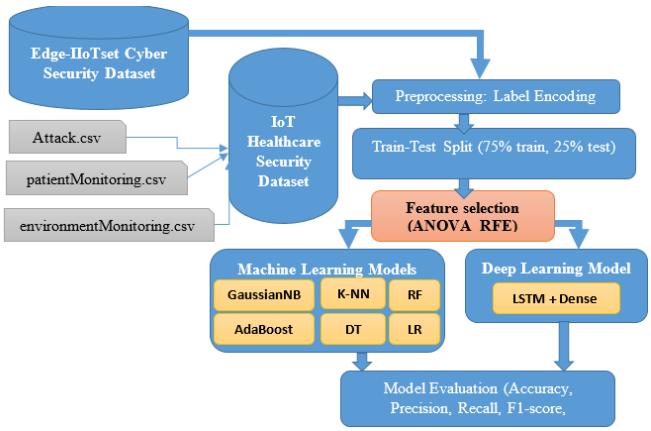


Figure 1. The proposed methodology

Initially, the dataset is acquired and preprocessed in order to get all columns in a numerical form. For the first dataset, three separate files are concatenated together to constitute the overall dataset. (The attack file, the patient monitoring information file, and the environment monitoring information file). For the next step, the label encoding algorithm is used. This step is essential to guarantee that all columns have numerical values so that they can be used in the training process.

The second step is the split of the dataset into a training set (75%) and a test set (25%). The training set will be used to train models while the test set will be used for the validation step.

For the next step, the proposed hybrid ANOVA-RFE feature selection algorithm will be applied in order to select the best subset of features (columns) of both datasets.

Many ML and DL models will be trained using the extracted features of the first and second datasets. The trained models are K-NN, LR, RF, DT, Gaussian Naïve Bayes GNB, AB, and the long-short Term Memory LSTM models. Many training scenarios are involved and the result models are evaluated using the test set and many performance evaluation metrics, including accuracy, precision, recall, F1-score and confusion matrix.

4.3 ML and DL models

Many ML algorithms are used for many security options as good models to make machines take decisions on critical problems (especially in the field of security attack detection and classification). In the current study, many ML and DL models are used, including RF, K-NN, GNB, LR, DT, AB, and LSTM [16].

K-NN is one of the common ML non-parametric algorithms predicting the category of data sample based on its K-nearest neighbors. The main hyperparameters values in K-NN is the K value that it's needed to be tuned. K-NN is a very easy and effective algorithm in case of small datasets (low dimensionality with low number of features) [17].

NB is a type of probabilistic algorithm that uses the Naïve Bayes theory. NB deals with features as they are independent of each other. This algorithm is fast and efficient with high-dimensional datasets [18].

Another non-parametric ML algorithm is the DT model by which the input space is divided into regions using some decision rules. The input data is split based on a specific feature and the idea of information gain. The split operation persists until the regions are pure or the stop condition is met. The main advantage of this algorithm is that it can process the numerical and categorical data. It's considered to be efficient in the case of classification and regression algorithms and for all data sizes [19].

RF, on the other hand, is an ensemble of many decision trees. It fuses those models to achieve the best performance and reduce overfitting. RF randomly selects a subset of features and then creates decision trees for each one. This mechanism reduces error and increases accuracy. RF is commonly used in both classification and regression models and is suitable for large and high-dimensional datasets (too many features) [20]. LR model is one of the linear models that are used to predict the sample into 0 or 1. Fitting a logistic function to the input is the main step in this algorithm. This algorithm is good in the case of binary classification on small and medium datasets.

Another ensemble model is the AdaBoost by which many weak models are fused together to build a stronger model. AdaBoost works to effectively reweight the data samples in each iteration in order to detect the misclassified data samples and trained a model using them. The final decision of this algorithm is based on the aggregation of predictions of all weak models. This algorithm is useful in the case of classification problems and is best suitable for small and medium datasets. It may cause some overfitting in case of high-dimensional datasets [17].

LSTM is a deep learning model that is mainly used for language and time series processing. It's considered as one type of recurrent neural network having a variate memory in order to memorize some connections and information about the data. This type of deep network can handle long-term dependencies by using a cell state that can be erased or updated based on the current input and the inputs of previous times [21].

4.4 Proposed ANOVA-RFE feature selection algorithm

This new algorithm is a hybrid algorithm that combines the benefit of both feature-based and wrapper-based methods. Since the recursive elimination algorithm eliminate the least essential K features of the dataset, it helps to avoid overfitting and improve the performance [22]. On the other hand, ANOVA can define the most essential features with the highest correlation to the target [23]. Combining these two advantages gives a very efficient feature selection model. In this study, an iterative algorithm of ANOVA-RFE is proposed. Details of this method are illustrated in Figure 2. The algorithm initially selects the best features using ANOVA

select Kbest algorithm with KANOVA=20. After that, the RFE algorithm works to select the best KRFE features throughout the ANOVA-based selected subset of features instead of looking up in the entire dataset. The RFE iterates until finding the best accuracy with the best combination of features.

In the proposed feature selection algorithm, the hyperparameters KANOVA and KRFE are chosen through a trial-and-error process to achieve the best accuracy with the best combination of features. First, the KANOVA is set to 20 to select the 20 best features with the highest correlation to the target using the ANOVA select Kbest algorithm. This value is chosen based on previous knowledge to ensure selecting a moderate number of features can lead to better classification performance and reduce overfitting. Second, the RFE algorithm works to select the best KRFE features throughout the ANOVA-based selected subset of features. The value of KRFE is selected based on the number of features in the ANOVA-based selected subset. The value of KRFE is varied from 5 to 20 and the one that give the best accuracy on the validation set is selected.

4.5 Performance evaluation metrics

To evaluate the trained models and the feature selection algorithms, many performance metrics can be used, including the following [24, 25].

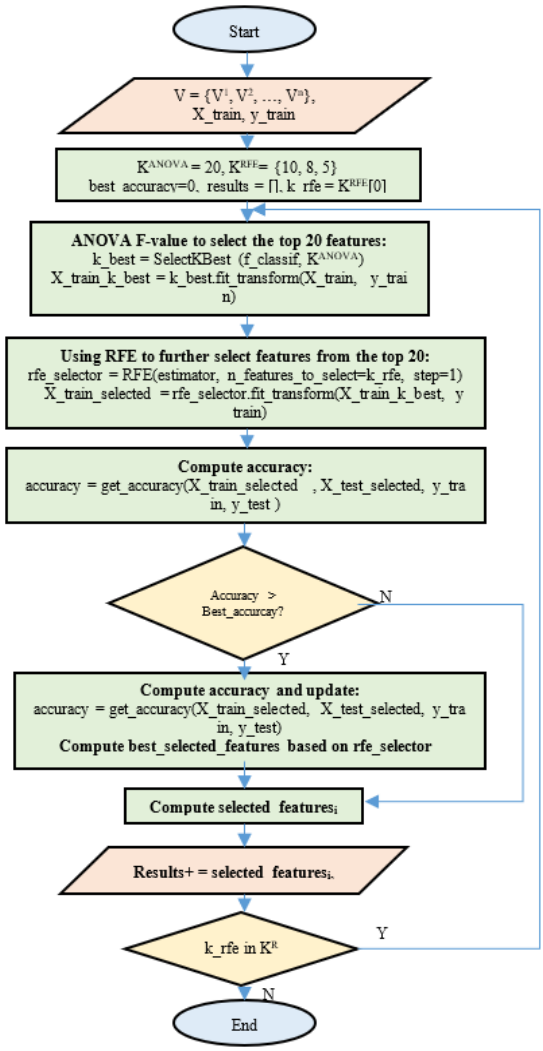


Figure 2. ANOVA-RFE proposed feature selection algorithm

Precision: the percentage of true positive test samples out of the entire predictive positives. It's given by Eq. (1). The accuracy of positive prediction is measured by this factor.

$$\text{Precision} = \text{true positives} / (\text{true positives} + \text{false positives}) \quad (1)$$

Recall: the percentage of true positive test samples out of the entire actual positives. Using recall, the ability of model to detect all positive cases can be measured. Recall calculation is shown in Eq. (2).

$$\text{Recall} = \text{true positives} / (\text{true positives} + \text{false negatives}) \quad (2)$$

F1-score: a mixture parameter that mixes the precision and recall together and is given as in Eq. (3).

$$\text{F1-score} = 2*\text{precision}*\text{recall} / (\text{precision} + \text{recall}) \quad (3)$$

Accuracy: the ratio of all true classified samples to the total number of samples.

Training time: to compare the training time before and after the selection step so that the benefit of using a feature selection model will appear.

Confusion matrix: a detailed results explaining the individual categories results (precision, recall and F1-socre). The confusion matrix plots the result between the true labels and the predicted labels for all categories.

5. RESULTS

In order to define the best model with the best training choice, the following training scenarios are proposed on the first dataset:

(1) Training ML and DL models without feature selection.

(2) Training ML and DL models with feature selection.

1) Feature Methods: Variance Threshold
2) Feature Methods: Select K-Best (K-Highest score features)
3) Feature Methods: Selecting best 10 features
4) Wrapper methods: Forward Selection
5) Feature-based and Wrapper-based hybrid method (ANOVA_RFE) non-iterative algorithm.
6) Feature-based and Wrapper-based hybrid method (ANOVA_RFE) iterative-based algorithm.
7) Score-level fusion of the best trained ML models (feature selection-based trained models) in which the scores of the best three trained models will be weighted and summed to configure a fused score. Then the final classification decision is made.

For the second dataset, the best feature selection method will be applied and compared to the original entire dataset.

5.1 Results of experiments applied without feature selection

5.1.1 Results on the first dataset

The experiments are first executed on the first dataset (binary classification problem). Table 1 shows the results of evaluating the trained ML and DL models using the test set of the first dataset.

Table 1 shows that K-NN, RF, DT, and AdaBoost algorithms achieve an accuracy of 100% besides all other performance metrics which are all 100%. The Next best model

is the LSTM model with a 99.94% score on all metrics. However, among the best-registered results, DT has the least computational time.

5.1.2 Results on the second dataset

For the second dataset, the experiments are repeated with the same parameters of all ML and DL models. Table 2 illustrates the results of evaluating the trained ML and DL models using the test set of the second dataset. For this dataset, the weighted average is only listed in Table 2 for each model since the number of categories is 15.

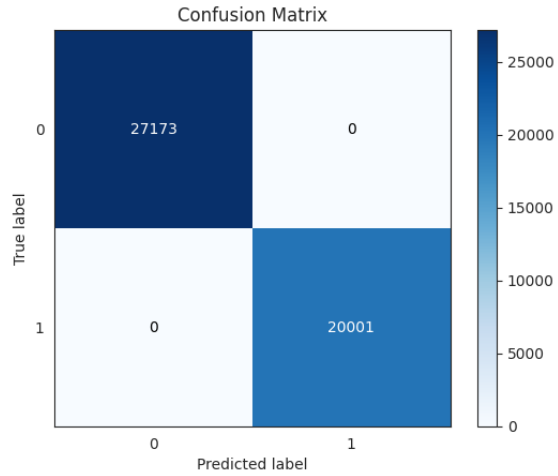
Table 2 shows that K-NN, RF, and DT algorithms achieve high performance with a transcendence of the RF model (accuracy of 99.99%). However, among the best-registered results, DT has the least computational time. Some ML models achieve a very low performance using the second dataset compared to the corresponding results of the first dataset. This is due to the fact that some ML algorithms are affected by the huge dataset size and the occurrence of multiple categories of the target class. Figure 3 shows the confusion matrix of the best-trained model using all features of the first and second datasets.

Table 1. Performance evaluation of the ML and DL models trained using the original first dataset (without feature selection)

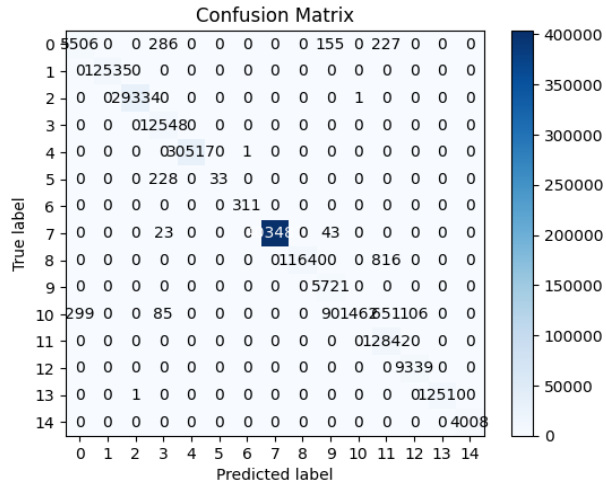
	Category	Accuracy%	Precision%	Recall%	F1-score%	Tr-Time
GNB	Normal	99.86	99.9	99.86	99.88	284 ms
	Attack		99.82	99.86	99.84	
	Weighted Average		99.86	99.86	99.86	
K-NN	Normal	100	100	100	100	1min 13s
	Attack		100	100	100	
	Weighted Average		100	100	100	
RF	Normal	100	100	100	100	8.61 s
	Attack		100	100	100	
	Weighted Average		100	100	100	
AB	Normal	100	100	100	100	14.1 s
	Attack		100	100	100	
	Weighted Average		100	100	100	
LR	Normal	98.91	98.85	99.26	99.05	2.69 s
	Attack		98.98	98.44	98.7	
	Weighted Average		98.91	98.91	98.91	
DT	Normal	100	100	100	100	464 ms
	Attack		100	100	100	
	Weighted Average		100	100	100	
LSTM Deep Model	Normal	99.94	99.93	99.97	99.95	10.2 s
	Attack		99.96	99.9	99.93	
	Weighted Average		99.94	99.94	99.94	

Table 2. Performance evaluation of the ML and DL models trained using the original second dataset (without feature selection)

	Accuracy%	Precision%	Recall%	F1-score%	Tr-Time
GNB	35.35	88.19	35.35	36.81	8.11 s
K-NN	94.58	94.22	94.58	93.75	1h 56min 48s
RF	99.99	99.98	99.81	99.89	3min 4s
AB	78.24	73.85	78.24	74.58	5min 7s
LR	75.02	61.69	75.02	66.06	3min 17s
DT	99.45	99.5	99.46	99.41	9.4 s
LSTM	87.79	88.91	87.8	87.4	102 s



(A) RF, DT and AdaBoost (best evaluated models of the first dataset)



(B) The RF model (best evaluated model of the second dataset)

Figure 3. Confusion matrixes of the trained models using all features of the first and second datasets: (A), (B)

5.2 Results of experiments applied with individual feature selection methods

In this part, the feature selection methods, including the feature-based and wrapper-based methods will be applied (each of which is a sperate training scenario), then the trained ML and DL models will be evaluated using the test set. Experiments will be applied to both the first and the second datasets.

The ANOVA-RFE method is chosen since it combines the benefits of two different feature selection algorithms. While ANOVA selects the k-best features with the highest correlation to the target feature, RFE eliminates the least significant features of the selected subset of features of ANOVA step. This hybrid feature selection method can effectively reduce the number of features and improve the classification performance. The ANOVA-RFE is chosen over PCA since PCA sometimes lead to a loss of interpretability The ANOVA-RFE is chosen over Lasso because the utilized datasets contain a large number of features, and Lasso may not be able to effectively select the best features. The ANOVA-RFE is chosen over ReliefF because ReliefF may not be suitable for datasets with a large number of features.

5.2.1 Results on the first dataset

The first feature selection method is the variance threshold with a threshold value of 0. The algorithm leads to 41 out of 51 features with a reduction rate of (19.6%). The result of evaluating models of this scenario is shown in Table 3.

Table 3 shows that the RF, AdaBoost and DT models remain the same performance (100%) even after dropping 19.6% of the columns. The DT model has the best computational time (311ms) which is much less than the original DT model trained using the entire columns.

The second feature selection method is the SelectKBest algorithm in which the K features with the highest degree of importance will be selected. In the current scenario, two experiments are involved; one using K=20 with a reduction rate of (60.78%), the another one using K=10 getting a reduction rate of (80.39%). The results of training ML and DL models using the selected features of "SelectKBest" algorithm are illustrated in Table 4.

Table 4 shows that the RF and AdaBoost models remain the same performance (100%) even after dropping 80.39% of the columns. The training time of all models has been significantly decreased (since the number of features is reduced). The most computational time enhancement is related to the K-NN classifier with almost (84%) time enhancement.

For the third feature selection method, the forward selection (wrapper methods) is used. In these methods, the sequential selection SFS method is used to select the most correlated features with the target (best 10 ones). The underlying model of SFS is the RF model. The results of this this feature selection method are shown in Table 5.

Table 3. Performance evaluation of the ML and DL models using the selected features of the first dataset using Variance threshold method

	Accuracy%	Precision%	Recall%	F1-score%	Tr-Time
GNB	99.86	99.86	99.86	99.86	167 ms
K-NN	99.97	99.97	99.97	99.97	1min 2s
RF	100	100	100	100	8.9 s
AB	100	100	100	100	13.2 s
LR	98.91	98.91	98.91	98.91	2.29 s
DT	100	100	100	100	311 ms
LSTM	99.97	99.98	99.98	99.98	7.15 s
Fusion Best three	99.97	99.98	99.98	99.98	-

Table 4. Performance evaluation of the ML and DL models using the selected features of the first dataset using the "SelectKBest" method

	Accuracy%		Precision%		Recall%		F1-score%		Tr-Time	
K	20	10	10	10	20	10	20	10	20	10
GNB	99.9	98.96	99.9	98.98	99.9	98.96	99.9	98.96	77.2 ms	54.5 ms
K-NN	99.99	99.9	99.99	99.99	99.99	99.99	99.99	99.99	40.6 s	6.47 s
RF	100	100	100	100	100	100	100	100	9.56 s	8.98 s
AB	100	100	100	100	100	100	100	100	9.65 s	6.09 s
LR	98.94	98.5	98.94	98.53	98.94	98.52	98.94	98.52	1.5 s	1.27 s
DT	100	99.99	100	99.99	100	99.99	100	99.99	216 ms	206 ms
LSTM	99.92	99.91	99.92	99.91	99.92	99.91	99.92	99.91	2min 23s	5.5 s
Fusion Best three	99.92	99.9	99.92	99.91	99.92	99.91	99.92	99.91	-	-

Table 5. Performance evaluation of the ML and DL models using the selected features of the first dataset using the forward elimination method

	Accuracy%	Precision%	Recall %	F1-score %	Tr-Time
GNB	88.75	90.95	88.75	88.4	39.6 ms
K-NN	100	100	100	100	2.96 s
RF	100	100	100	100	5.87 s
AB	100	100	100	100	5.49 s
LR	99.59	99.6	99.6	99.6	2.52 s
DT	100	100	100	100	120 ms
LSTM	99.97	99.97	99.97	99.97	7.15 s
Fusion	100	100	100	100	-

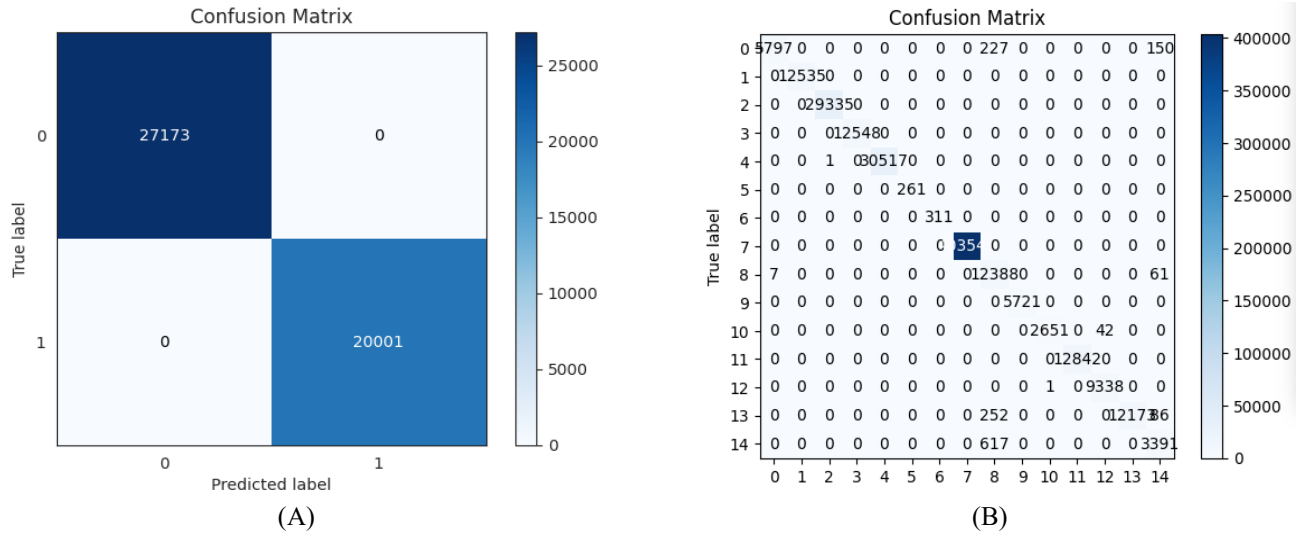
F: Forward selection, **B**: Backward selection.

Table 6. Performance evaluation of the ML and DL models using the selected features of the first dataset using the ANOVA-RFE method

	Accuracy%	Precision%	Recall%	F1-score%	Tr-Time
GNB	99.27	99.28	99.28	99.27	46.2 ms
K-NN	99.97	99.97	99.97	99.97	2.71 s
RF	100	100	100	100	5.9 s
AB	100	100	100	100	6.25 s
LR	99.14	99.14	99.14	99.14	1.21 s
DT	100	100	100	100	127 ms
LSTM	99.97	99.97	99.97	99.97	4.7 s
Fusion	100	100	100	100	-

Table 7. Performance evaluation of the ML and DL models using the selected features of the second dataset using the ANOVA-RFE method

	Accuracy%	Precision%	Recall%	F1-score%	Tr-Time
GNB	53.039	90.47	53.04	61.76	1.21 s
K-NN	99.54	99.56	99.54	99.52	1min 13s
RF	99.84	99.84	99.84	99.84	2min 26s
AB	78.24	73.85	78.24	74.58	2min 27s
LR	78.03	63.94	78.03	69.54	2min 3s
DT	99.68	99.71	99.68	99.69	4.33 s
LSTM	98.61	98.7	98.62	98.44	4.65 s
Fusion	99.68	99.71	99.68	99.68	-



(A) (B)

Figure 4. Confusion matrixes of the trained RF models using the ANOVA-RFE selected features of the first (A) and second (B) datasets

The fourth selection method is the hybrid proposed algorithm. For this algorithm, the experiments are first applied using specific parameters (number of selected features using ANOVA=30, number of features to select using RFE is 10). The results of this scenario are shown in Table 6.

The experiments on the first dataset show that the fusion has no enhancement effect. Besides, all feature selection algorithms lead to reducing the computational time with reserving or a very small decrease in the performance.

5.2.2 Iterating the ANOVA-RFE algorithm on the first dataset

The proposed ANOVA-RFE feature selection algorithm is dynamically performed to determine the optimal number of selected features that maintain the accuracy of the best model and reduce the training time. Three different numbers of selected features are tried and the accuracy is computed for each iteration. The results on the first dataset show that the algorithm selects the best 10, then the best 8 then the best 5 features with preserving the same performance 100% of the RF model with a training time of only 4.2 seconds (51.22%-time enhancement compared to the original dataset (8.61 s)) and a reduction rate of 90.19%. The list of the best-selected features is: ['frame.time_relative' 'ip.src' 'ip.dst' 'tcp.time_delta' 'tcp.hdr_len'].

5.2.3 Results on the second dataset

Algorithm (hybrid ANOVA-RFE). The algorithm is applied in the same scenario as the first dataset. The result of evaluating all ML and DL models trained using the selected features of ANOVA-RFE is illustrated in Table 7.

Table 7 illustrates that the best performance is related to the RF algorithm with 99.84% accuracy. The least computational time with high performance is the DT algorithm with 4.33 s and 99.68% accuracy.

5.2.4 Iterating the ANOVA-RFE algorithm on the second dataset

In this part, the ANOVA-RFE is iterated in order to get the

best combination of features with the best performance.

The results on the second dataset show that the algorithms selected the best 10, then the best 8 then the best 5 features with 99.98%, 99.98% and 99.96% for the three cases, respectively. These results indicate that the best case is the case of 5 selected features since the performance decreased only by 0.02% while the number of features is decreased by 91.93%. The training time of RF model using the entire features is almost 3 minutes (180000ms) while it is decreased to only 158.32ms by using the selected 5 features (i.e., time enhancement of 99.9%). The selected five features are: 'frame.time' 'tcp.options' 'dns.qry.name.len' 'mqtt.conack.flags' and 'mqtt.topic'.

Figure 4 shows the confusion matrix of the best-trained model using the 5 selected features of the first and second datasets.

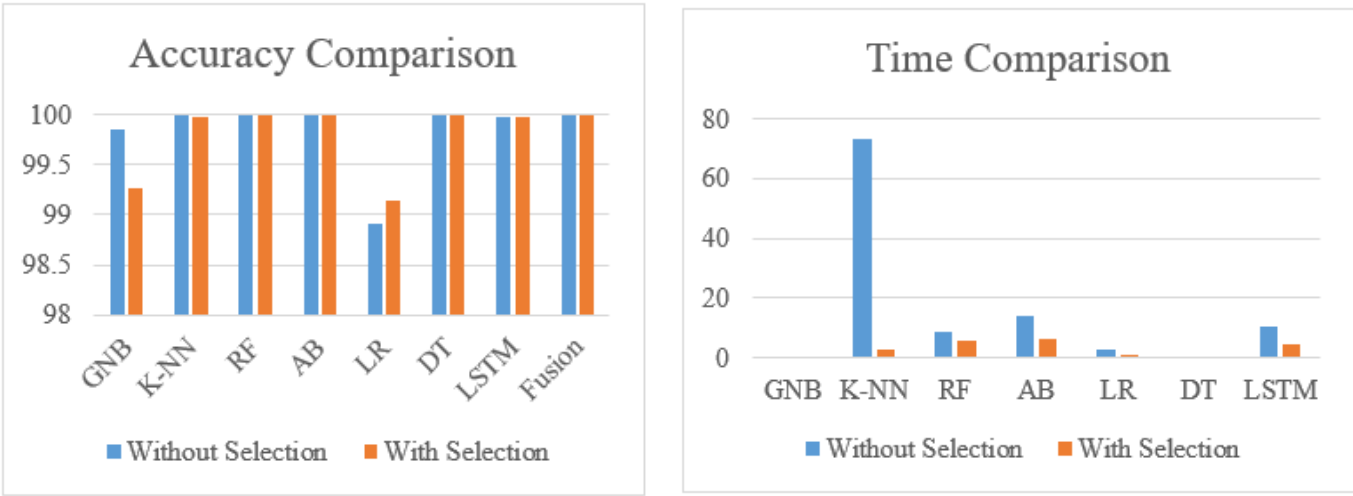
6. DISCUSSION

In order to demonstrate the effectiveness of the proposed feature selection algorithm, the evaluation results of trained models that were based on both the entire set of features and the selected set of features will be compared. Figure 5 illustrates this comparison.

Similarly, for the second dataset, Figure 6 illustrates the same comparison., Figure 6 illustrates the same comparison.

Figure 5(A) illustrates that the performance in the case of a feature reduction rate of 90.19% of the original features is either the same as the original case (using entire features) or has a higher accuracy although the high reduction rate. In the case of the GNB model, the accuracy after selection is minimized by 0.59%. For the second dataset scenarios, the training using the selected subset of features always leads to higher accuracy for all ML and DL models.

Figure 5(B) and Figure 6(B) prove that the computational time of the training process using the selected features is less than the computational time of the trained models using the entire feature set for both datasets. The reducing rate is noticed especially for the K-NN model since it is affected by the high dimensionality of data.

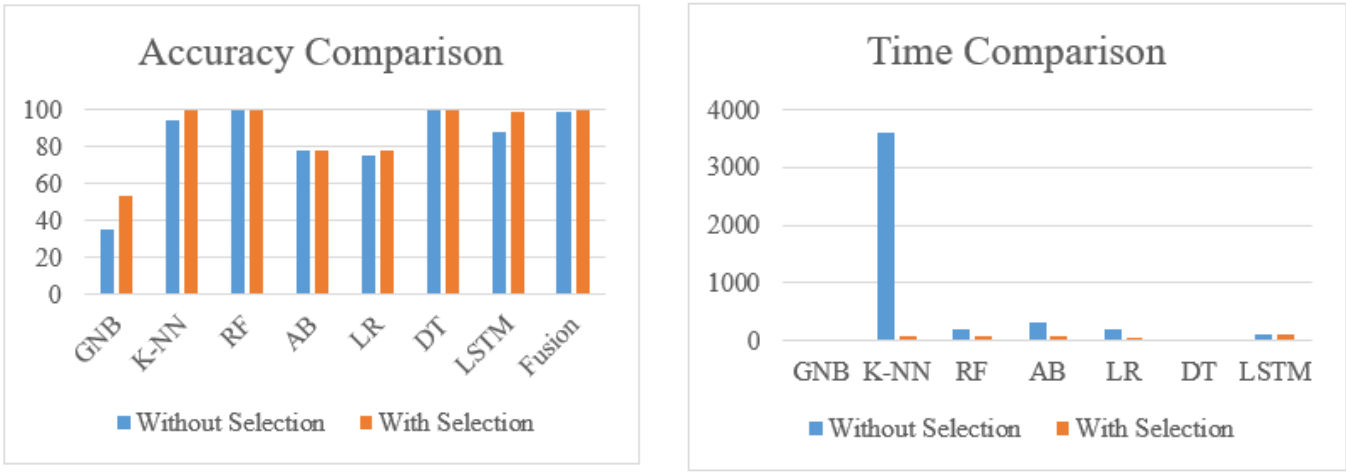


(A) (B)

Figure 5. Accuracy and time-based comparison between using all and selected features of the first dataset (A) Accuracy, (B) Time

Table 8. A comparison between the current study and related work

Researcher	Methodologies	Dataset	Results	Notes
Vinayakumar et al. [5]	Deep learning architecture, two-tier environment	DS1-V1, DS2-V2	99.2%, 89.9%	Limited to binary classification (Normal or attack)
Popoola et al. [6]	LAE-BiLSTM architecture, binary classification	BotIoT	91.89%	Limited to binary classification (Normal or attack)
Hussain et al. [7]	IoT-Flock, ML algorithms (RF, NB, AB, LR, DT)	Not specified	99.51% (RF)	Limited to healthcare IoT environment (due to the utilized dataset type)
Ferrag et al. [8]	Feature reduction, ML classifiers (RF, SVM, DNN)	Edge-IIoTset	94.67% (DNN)	They created a new dataset Edge-IIoTset Cyber Security
Bahadoripour et al. [10]	Multi-modal deep learning model	Secure Water Treatment	98% (F1-score)	Binary classification problem (Normal or attack)
Kumar and Sharma [11]	CNN-based model	Not specified	AUC of 0.993	Binary classification problem (outliers or inliers)
Rashid et al. [12]	Federated Learning (FL)	Edge-IIoTset	93.92%	FL method reliability is low, FL has limitations (time and accuracy)
Hnamte and Hussain [13]	Hybrid deep learning model (BiLSTM and CNN)	CICIDS2018, Edge_IIoT	100%, 99.64%	No feature selection (High dimensionality)
Alzahrani R. and Alzahrani A. [14]	Statistical method (K-NN, cumulative sum, exponentially weighted average)	Bot-IoT	99%	DDoS attack detection only (one type of attacks)
Cuurent Study	RF, LR, DT, K-NN, GNB, AdaBoost, LSTM, AVONA-RFE	Edge-IIoTset	Cuurent Study	RF, LR, DT, K-NN, GNB, AdaBoost, LSTM, AVONA-RFE



(A) (B)

Figure 6. Accuracy and time-based comparison between using all and selected features of the second dataset: (A) Accuracy, (B) Time

All models in the second dataset scenarios performed better in case of feature selection as shown in Figure 6(A). This is due to the fact that some features in the original second dataset are not just redundant but also make some bias to specific classification resulting in classification errors. The characteristics of the data and the problem may be more complex than in the first dataset scenario (i.e., the second dataset may contain some outliers or noisy data so the feature selection algorithm can reduce these outliers by removing features with too much noise and improve the performance).

6.1 Comparing the current study with related work

In order to specify the importance and efficiency of the current study, Table 8 includes a comparison between the current study and related work.

6.2 Limitations and future work

The limitations of the current study can be concluded by two main issues; the first one is that the proposed ANOVA-RFE algorithm needs more experiments using different dataset's size and challenges, while the second issue is the using of built-in models without creating new ones. The proposed feature selection algorithm needs to be evaluated using new models. Future work can focus on applying the ANOVA-RFE on different datasets with different challenges. Besides, the future studies can focus on evaluating the ANOVA-RFE algorithm on new DL models.

7. CONCLUSIONS

In the current study, a new security attack detection system based on machine learning, deep learning and feature selection was proposed. A new hybrid ANOVA-RFE feature selection algorithm is designed and implemented. Experiments were applied on two different datasets and under different training scenarios. The first dataset contains 52 columns and 188694 records, while the second dataset includes 63 columns and 2219200 records. Tests prove the following:

(1) The AVOVA-RFE feature selection algorithm dynamically selected the best 5 features of the first dataset with an accuracy of 100% of the RF model and a computational time enhancement of 51.22% and a reduction rate of 90.19%.

(2) The AVOVA-RFE feature selection algorithm dynamically selected the best 5 features of the first dataset with an accuracy of 99.96% of the RF model and a computational time enhancement of 99.91% and a reduced rate of 91.93%.

(3) The main significant of this study is that the proposed ANOVA-RFE algorithm improved the accuracy and reduced the training time of machine learning models for security attack detection. Moreover, the proposed ANOVA-RFE can be used in other security applications like ransomware detection, malware detection, etc.

(4) Comparing our study with the previous ones in the same field proves the efficiency and transcendence of the proposed feature selection algorithm.

Comparing our study to previous ones proves that the current study outperformed the related work, However, it is limited to the used datasets, so future studies can focus on applying different feature selection algorithms on different datasets. However, this depends on the specific utilized datasets, models and performance metrics used in this study. Future studies can combine the benefits of feature-based and wrapper-based methods, such as the mutual information-based feature selection algorithm, the genetic algorithm-based feature selection algorithms, auto encoder-based feature selection algorithm or the deep belief network-based feature selection algorithm. Future work could investigate the use of more advanced machine learning and deep learning models, such as reinforcement learning-based models, graph neural networks, or attention-based models. Using of datasets with different level of outliers, noise, class imbalance, etc. could be a good choice for future works.

REFERENCES

[1] Abdullahi, M., Baashar, Y., Alhussian, H., Alwadain, A., Aziz, N., Capretz, L.F., Abdulkadir, S.J. (2022). Detecting cybersecurity attacks in internet of things using artificial intelligence methods: A systematic literature review. Electronics, 11(2): 198. https://doi.org/10.3390/electronics11020198

[2] Kebande, V.R. (2022). Industrial internet of things (IIoT)

forensics: The forgotten concept in the race towards industry 4.0. Forensic Science International: Reports, 5: 100257. https://doi.org/10.1016/j.fsir.2022.100257

[3] Kadhim, I.B., Khaleel, M.F., Mahmood, Z.S., Coran, A.N.N. (2022). Reinforcement learning for speech recognition using recurrent neural networks. In 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), Ravet, India, pp. 1-5. https://doi.org/10.1109/ASIANCON55314.2022.9908930

[4] Mohammed, A.B., Chaari Fourati, L., Fakhrudeen, A. M. (2022). A comparative study of attribute selection algorithms on intrusion detection system in UAVs: A case study of UKM-IDS20 dataset. In International Conference on Risks and Security of Internet and Systems, Sousse, Tunisia, pp. 34-46. https://doi.org/10.1007/978-3-031-31108-6_3

[5] Vinayakumar, R., Alazab, M., Srinivasan, S., Pham, Q.V., Padannayil, S.K., Simran, K. (2020). A visualized botnet detection system based deep learning for the internet of things networks of smart cities. IEEE Transactions on Industry Applications, 56(4): 4436-4456. https://doi.org/10.1109/tia.2020.2971952

[6] Popoola, S.I., Adebisi, B., Hammoudeh, M., Gui, G., Gacanin, H. (2020). Hybrid deep learning for botnet attack detection in the internet-of-things networks. IEEE Internet of Things Journal, 8(6): 4944-4956. https://doi.org/10.1109/JIOT.2020.3034156

[7] Hussain, F., Abbas, S.G., Shah, G.A., Pires, I.M., Fayyaz, U. U., Shahzad, F., Zdravevski, E. (2021). A framework for malicious traffic detection in IoT healthcare environment. Sensors, 21(9): 3025. https://doi.org/10.3390/s21093025

[8] Ferrag, M.A., Friha, O., Hamouda, D., Maglaras, L., Janicke, H. (2022). Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning. IEEE Access, 10: 40281-40306. https://doi.org/10.1109/ACCESS.2022.3165809

[9] Tang, Z., Hu, H., Xu, C. (2022). A federated learning method for network intrusion detection. Concurrency and Computation: Practice and Experience, 34(10): e6812. https://doi.org/10.1002/cpe.6812

[10] Bahadoripour, S., MacDonald, E., Karimipour, H. (2023). A deep multi-modal cyber-attack detection in industrial control systems. In 2023 IEEE International Conference on Industrial Technology (ICIT), Orlando, FL, USA, pp. 1-6. https://doi.org/10.48550/arXiv.2304.01440

[11] Hussein, S.F., Dallalbashi, Z.E., Mohammed, A.B. (2023). Anomaly detection in internet of medical things with artificial intelligence. Eastern-European Journal of Enterprise Technologies, 121(4): 56-62. https://doi.org/10.15587/1729-4061.2023.274575

[12] Rashid, M.M., Khan, S.U., Eusufzai, F., Redwan, M.A., Sabuj, S.R., Elsharief, M. (2023). A federated learning-based approach for improving intrusion detection in industrial internet of things networks. Network, 3(1): 158-179. https://doi.org/10.3390/network3010008

[13] Hnamte, V., Hussain, J. (2023). DCNNBiLSTM: An efficient hybrid deep learning-based intrusion detection system. Telematics and Informatics Reports, 10: 100053. https://doi.org/10.1016/j.teler.2023.100053

[14] Alzahrani, R.J., Alzahrani, A. (2023). A novel multi algorithm approach to identify network anomalies in the IoT using Fog computing and a model to distinguish between IoT and Non-IoT devices. Journal of Sensor and Actuator Networks, 12(2): 19. https://doi.org/10.3390/jsan12020019

[15] Hussain, F., Abbas, S.G., Shah, G.A., et al. (2021). IoT Healthcare Security Dataset. IEEE Dataport. https://dx.doi.org/10.21227/9w13-2t13

[16] Gowtham, M., Pramod, H.B. (2021). Semantic query-featured ensemble learning model for SQL-injection attack detection in IoT-ecosystems. IEEE Transactions on Reliability, 71(2): 1057-1074. https://doi.org/10.1109/tr.2021.3124331

[17] Liu, Y., Yang, M., Wang, Y., Li, Y., Xiong, T., Li, A. (2022). Applying machine learning algorithms to predict default probability in the online credit market: Evidence from China. International Review of Financial Analysis, 79: 101971. https://doi.org/10.1016/j.irfa.2021.101971

[18] Metipatil, P., Bhuvaneshwari, P., Basha, S.M., Patil, S.S. (2023). An efficient framework for classifying cancer diseases using ensemble machine learning over cancer gene expression and sequence-based protein interactions. In 2023 2nd International Conference for Innovation in Technology (INOCON), Bangalore, India, pp. 1-8. https://doi.org/10.1109/inocon57975.2023.10101354

[19] Mahmood, Z.S., Coran, A.N.N., Kamal, A.E., Noori, A.B. (2021). Dynamic spectrum sharing is the best way to modify spectrum resources. In 2021 Asian Conference on Innovation in Technology (ASIANCON), Pune, India, pp. 1-5. https://doi.org/10.1109/ASIANCON51346.2021.9544912

[20] Sarmas, E., Spiliotis, E., Dimitropoulos, N., Marinakis, V., Doukas, H. (2023). Estimating the energy savings of energy efficiency actions with ensemble machine learning models. Applied Sciences, 13(4): 2749. https://doi.org/10.3390/app13042749

[21] Haq, M.A., Ahmed, A., Khan, I., Gyani, J., Mohamed, A., Attia, E.A., Mangan, P., Pandi, D. (2022). Analysis of environmental factors using AI and ML methods. Scientific Reports, 12(1): 13267. https://doi.org/10.1038/s41598-022-16665-7

[22] Chen, R.C., Manongga, W.E., Dewi, C. (2022). Recursive feature elimination for improving learning points on hand-sign recognition. Future Internet, 14(12): 352. https://doi.org/10.3390/fi14120352

[23] Pathan, M.S., Nag, A., Pathan, M.M., Dev, S. (2022). Analyzing the impact of feature selection on the accuracy of heart disease prediction. Healthcare Analytics, 2: 100060. https://doi.org/10.1016/j.health.2022.100060

[24] Kulkarni, A., Chong, D., Batarseh, F.A. (2020). Foundations of data imbalance and solutions for a data democracy. Data Democracy, pp. 83-106. https://doi.org/10.1016/B978-0-12-818366-3.00005-8

[25] Anand, M., Velu, A., Whig, P. (2022). Prediction of loan behaviour with machine learning models for secure banking. Journal of Computer Science and Engineering (JCSE), 3(1): 1-13. https://doi.org/10.36596/jcse.v3i1.237