An Efficient Anomaly Detection System in IoT Edge using Chi Square-Improved Particle Swarm Optimization Feature Selection with Ensemble classifiers

J Manokaran[1], Vairavel Gurusamy[2], Osamah Khalaf[3], Sameer Algburi[4] and Habib Hamam[5]

[1] *Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai, 603 203, India*
[2] *Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Tiruchirappalli, India*
[3] *Department of Solar, Al-Nahrain Research Center for Renewable Energy, Al-Nahrain University, Baghdad, Iraq*
[4] *College of Engineering Technology, Al-Kitab University, Kirkuk, Iraq*
[5] *Faculty of Engineering, University of Moncton, NB E1A3 E9, Canada*

E-mail address: manoraj3@gmail.com, vairavelg@protonmail.com, usama81818@nahrainuniv.edu.iq, sameer.algburi@uoalkitab.edu.iq, habib.hamam@umoncton.ca

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: Anomaly detection using machine learning (ML) algorithms is the key research theme in the modern digital era. Though the recent ML-based anomaly detection models have better detecting ability, the vast volume of data and its multi-dimensionality limit their ability with less accuracy, a low detection rate, and high learning complexity. This paper aims to enhance the performance of anomaly detection by combining various optimized ensemble learning algorithms, such as random forest (RF), extreme gradient boosting (XG Boost), adaptive boosting (Ada Boost), and light gradient boosting machine (LGBM), with a new hybrid feature selection approach. An evolved version of particle swarm optimization (IPSO) is initially developed, which integrates the elimination and opposition-based learning approaches to enhance PSO and then hybridizes it with the Chi-square method (Chi-IPSO). The developed model is evaluated using two standard datasets: UNSW NB 15, and CICIDS 2017. The research results show that the RF algorithm with Chi-IPSO performs better with an accuracy of 94.58% for the UNSW NB 15, and 99.70% for the CICIDS 2017. Several assessment measures, including F-score, MCC value, accuracy, precision, and recall, are used to highlight the outcome analysis of the suggested model. The results clearly show that the created model performs better than other modern approaches.

Keywords: High-dimensional data, Chi Square, Anomaly detection, Ensemble learning,IPSO, Feature selection, IoT security.

1. INTRODUCTION

The Internet of Things is one of the dominant communication paradigms that connects the physical and virtual worlds. It is revolutionizing the way people live and work. It extends to various domains, including smart cities, smart healthcare, smart banking, intelligent surveillance, and Industry 4.0 [1]. The IoT has a significant financial and community impact on people's lives. According to the IHS Markit report, the Internet will interlink over 100 billion IoT gadgets by 2030 (www.ihsmarkit.com). Since the IoT has a large volume and heterogeneity of devices, and is resource constrained, it has more security concerns than other sectors [2]. According to Kaspersky's statistics report, more than 15.37% of globally used Internet devices have experienced at least one attack. The WEF report 2023 identifies cyber-attacks as one of the top five sources of significant global risk [3].

In IoT network security, intrusion detection system (IDS) is critical for sensing and identifying unauthorized access to compute resources and networks. There are two categories of IDS: signature-based and anomaly-based system [4]. In the earlier one, the data pattern was already stored in the database and matched with the input sequences for detecting anomalies. This method is

E-mail:

simple and accurate, but it is not able to detect the modern attacks. Generating patterns manually for each attack is time-consuming and human-dependent, which is another limitation. The anomaly-based detection system examines network traffic and compares it to an ideal patterns to detect anomalies. The significant advantage of this method is that it can predict new attacks. Among the two types, we focus on anomaly-based detection methods.

Intrusion detection is a network analytic procedure in which ML algorithms are applied automatically to identify features of user's abnormal or regular activity [5]. Recent ML algorithms have a high false alarm rate, which makes it tough to detect new types of attacks accurately. Furthermore, different optimization and soft computing techniques have been used to improve anomaly detection performance. Combining several ML algorithms with weak predictive results generated through various projections of data, ensemble learning (EL) methods achieve better performance than any constituent method by combining results with different voting mechanisms [6]. Tama et al. conducted a structured mapping study of numerous EL algorithms used for abnormality detection. The authors concluded that the EL algorithms have better prediction accuracy than the single-learning algorithms [7].

The huge volume and diversification of attacks in the open IoT edge environment make anomaly detection more complex. The multi-dimensional nature of the data reduces the performance of the IDS, increasing the computational and model-building time. Feature selection (FS) plays an energetic role in EL algorithms, expanding detection accuracy of the learning algorithm. FS techniques are categorized as filters, wrappers, and embedded methods [8]. Filter methods are rapid and less computationally expensive, whereas the wrapper method exhibits better performance.

In our proposed system, we have combined the benefits of filter and wrapper methods using the Chi-IPSO algorithm for feature selection. To explore the entire search space and evaluate possible attribute subsets, the Chi-IPSO algorithm is used as a search algorithm. The model is trained with EL algorithms such as RF, LGBM, Ada Boosting, and XG Boosting. Additionally, we enhance the performance of the EL algorithms by tuning their parameters with the PSO algorithm. Another suggestive contribution of this study is that all the calculations occur at the edge of the device corner, which will increase the security and performance of the model. Figure 1 displays the proposed framework for edge computing.

We aim to create an efficient ADS for edge computing using an optimized EL algorithm. The suggested approach will produce the best classification accuracy with fewer features and less time. The key contribution of this study is as follows:

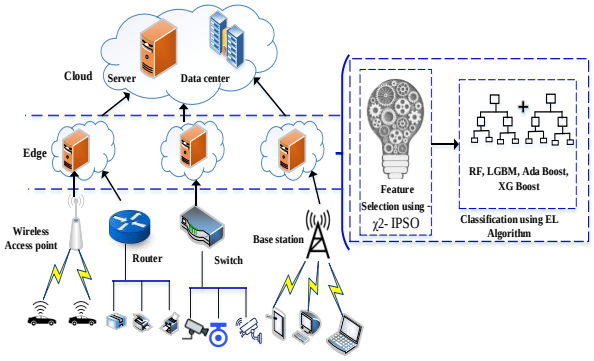


Figure 1. Anomaly detection at IoT Edge

- The PSO algorithm has been improved using elimination and opposition-based learning approaches and hybridized with the chi square method (Chi-IPSO).
- The proposed Chi-IPSO algorithm is used for the selection of optimum feature in anomaly detection model.
- Then we use different PSO optimized EL algorithms such as RF, Ada Boosting, LGBM, and XG Boosting algorithms for network anomaly detection in IoT scenario.
- The developed model of the RF algorithm with Chi-IPSO feature selection has been compared with contemporary methods using accuracy, F1 score, and MCC value.

The remaining paper is ordered as follows: Section. 2 offers recent similar studies about the use of optimization for efficient IDS; Section 3 provides the proposed ADS and its parameters; and Section. 4 presents the experimental setup of the developed system and results. Section. 5 provides the conclusions and further works.

2. LITERATURE REVIEW

According to the IDC cyber-security spending guide 2019, expenditures on security may exceed 133.8 billion dollars in 2023. In recent years, ML-based IDS has been a central research area worldwide, and several models have been established to detect anomalies in IoT.

Qusyairi et al. [9] improved the performance of an IDS using Spearman's rank coefficient-based feature selection and EL algorithms. Logistics regression (LR), decision tree (DT), and LGBM algorithms are chosen as the base classifiers and combined to attain 98.80% accuracy for the CICIDS 2018 dataset. The author

signifies that the classifier's hyper-parameters will be tuned and extended to a DL algorithm for large data prediction in future work. Albulayhi et al. [10] modeled an IDS using ML algorithms and novel feature selection algorithms. Optimum feature sets are selected using the set theory concept for filter-based FS approaches (Information Gain (IG) and Gain Ratio (GR)). The author concluded that for the IoTIDS 20 dataset, the developed algorithm attained 99.70 % accuracy. The implementation of the proposed algorithm on edge devices is suggested for future work.

Abdullah et al. [11] enhanced the performance of ADS using FS techniques and EL algorithms. The dataset was divided into multiple small sets, and IG feature selection was applied for each dataset. The RF, J48, and partial DT algorithms are applied for classification. The author indicates that the Ada boosting algorithm can be used for FS and classification in the future. Mhawi et al. created an IDS using hybrid FS method and voting ensemble concept [12]. Correlation feature selection coupled with forest panelized attributes is used to select the attributes. The K-nearest neighbours (KNN), support vector machine (SVM), RF, and naive bayes (NB) algorithms are applied as base algorithms and fused with majority voting concepts. The developed model achieved 99.70 % accuracy for the CICIDS 2017 dataset. In the future, more recent datasets may be used for evaluation.

Rahman et al. [13] invented a novel FS technique for an IoT attack detection system. The SVM, NB, and C4.5 algorithms are applied as base ML algorithms and merged using artificial neural networks (ANN). The proposed method achieved 99.90 % accuracy for the Aegean Wi-Fi intrusion dataset. Alghanam et al. [14] developed an IDS using improved pigeon-inspired optimization (PIO) feature selection with an EL algorithm. Local search combined with the PIO technique further enhances the selection of optimum features. The outcomes reveal that the advised method works better than other NIDS strategies published in recent years.

Baijnath et al. [15] compared the detection ability of various ML algorithms with FS techniques for IDS systems. The LR, RF, KNN, Ada Boost, DT, Gr Boost, NB, XG Boost, and SVM are the learning algorithms used. Chi-square and IG techniques are used for attribute selection. RF with a feature selection algorithm achieved a precision value of 99.12 %. Ebrima et al. [16] produced an IDS using lightweight hybrid FS techniques and EL algorithms. Initially, genetic search and rule-based feature selection techniques are combined, and then the expectation-maximization SVM, k-means, and DBSCAN algorithms are mixed using the majority voting principle. The proposed system has attained 99.70 % accuracy for the KDD Test-21 dataset.

Saikat et al. [17] performed a comparative analysis of NIDS using ensemble feature selection and ML algorithms. Nine prominent feature selection techniques (Recursive Feature Elimination (RFE), ANOVA, Chi-square, LASSO, RF, Mutual information, Pearson, LR, and SFPR) are combined using the majority voting principle and applied to the ensemble classifier. The developed algorithm attained an accuracy of 99.30 % and a 0.5 % error rate for the CICIDS 2017 datasets. In [18], Hui et al. created a new IDS using the PSO-XG Boost algorithm. The PSO method is used to optimize the XG Boost algorithm's settings. For the multi-classification, the parameter-optimized ensemble learning method has a high detection rate in comparison to the base ensemble learning algorithm.

From the related work, we observe that optimum feature selection will increase the performance of IDS, which is the aim of our work. For the critical IoT scenarios, the performance of a single FS technique is not the appropriate approach, so we have combined the filter and wrapper methods in our developed model.

3. PROPOSED METHODOLOGY

This research aims to create an effective ADS using the Chi-IPSO-based FS and optimized EL algorithms, which is shown in Figure 2. Initially, the input data are gathered from the online benchmark datasets: UNSW NB-15, and CICIDS 2017. Data scrubbing, encoding, and normalization are performed in the pre-processing stage. Optimum features are chosen by a hybrid of Chi square feature selection and the IPSO algorithm. The dataset is split 80/20 between training and testing data using the Hold-out method. Various EL algorithms like RF, XG Boost, Ada Boost, and LGBM algorithms are used to train and test the data. The final model's performance is assessed using standard performance metrics.

A. Data Collection and analysis

The nature of the information applied to train an EL algorithm significantly impacts the model's performance. The data must be cleansed for further analysis because it includes critical features about the problem domain. Here, our suggested approach is trained and tested on two popular datasets: UNSW NB [19] and CICIDS 2017 [20]. The UNSW-NB15 dataset focuses on network intrusions. It comprises nine attacks, including denial-of-service attacks, worms, backdoors, and fuzzes. The testing set has only 82,332 entries, while the training set contains 175,341 records. CICIDS 2017 is one of the latest and largest datasets used for IDS. The dataset has 79 features, 15 class samples, and seven attack categories. The dissemination of data is tabulated in Table 1.

TABLE I. DATA DISTRIBUTION FOR OUR MODEL ANALYSIS

Traffic Labels	UNSW NB 15		CICIDS 2017	
	Training	*Testing*	*Training*	*Testing*
Normal	56000	37000	318087	136219
Attack	119341	45322	78217	33626

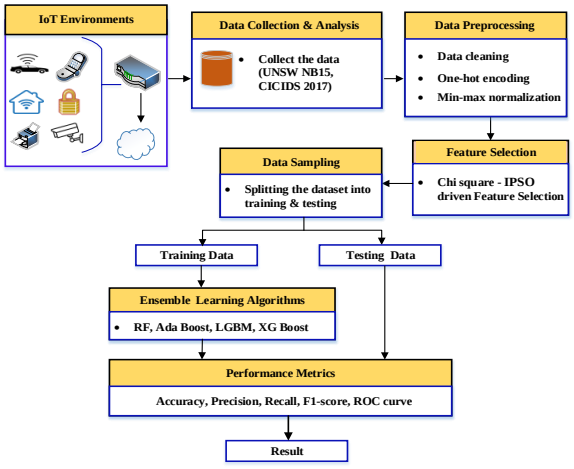


Figure 2. Proposed anomaly detection framework

B. *Data Preprocessing*

Pre-processing has the ability to accelerate ML algorithm. Three steps are taken in the preparation stage: filtering, encoding, and normalization. In the filtering phase, unwanted and null information's are changed by a threshold value. The categorical samples are changed into numeric samples in the data encoding stage by One-Hot encoding. Further, the data are normalized using Min-Max normalization techniques. Finally, the conversion of data is calculated using Equation 1.

$$\alpha_{new} = \frac{\alpha_i - \alpha_{\min}}{\alpha_{\max} - \alpha_{\min}} \qquad (1)$$

C. *Feature Selection (FS)*

FS is essential in model design, particularly in multi-dimensional IoT environments. Optimum FS can enhance the model's detection ability and decrease the model's training time. Compared to single-feature selection techniques, hybrid techniques have better results. In our proposed system, we use a hybrid of the Chi-square technique and the IPSO algorithm, which are explained below.

1. *Chi-Square technique*

The features in the training database are ranked using the χ^2 technique to determine which features are the most discriminative and have the highest detection accuracy [21]. considering the attributes t_i and the class label c_j the χ^2 is defined as follows,

$$\chi^2(t_i, c_j) = \frac{N(F_1 Z - Y F_2)^2}{(F_1 + F_2)(F_1 + Z)(Z + F_2)(Y + Z)} \qquad (2)$$

where F_1, F_2 are the occurrence of attributes t_i and class label c_j in the dataset. Y is the occurrence of c_j appearing without t_i. Z is the occurrence of neither c_j nor t_i appearing together in the dataset. N is the total number of samples in the dataset. Top 16 features are selected using chi-square techniques with feature score for different datasets that are shown in Figures 3-4.

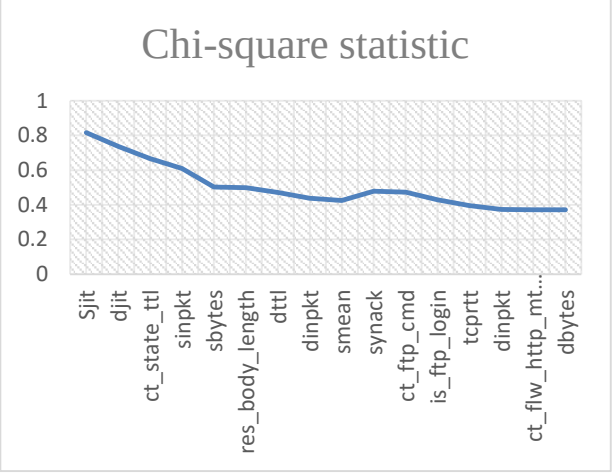


Figure 3. Chosen features for UNSW NB 15 dataset

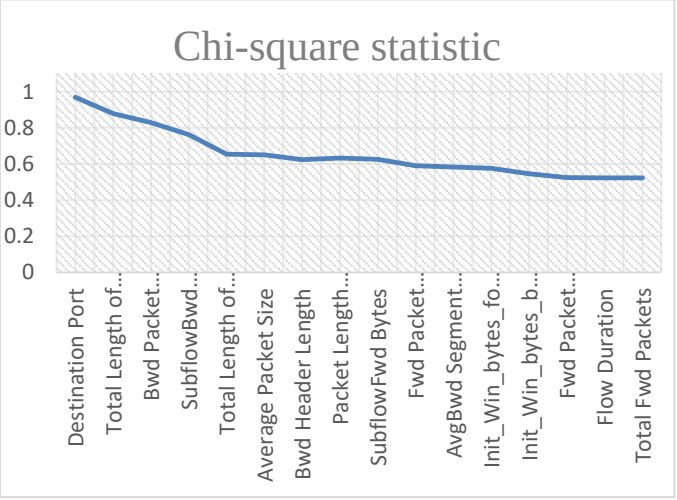


Figure 4. Chosen features for CICIDS 2017 dataset

2. Particle Swarm Optimization algorithm (PSO)

PSO is a well-known intelligent searching techniques developed in 1995 based on the principle of bird predation. The PSO approach is zero-order and derivative-free. This implies that it is not dependent on gradients, making it applicable to a wide range of situations, including multi-modal and discontinuous problems. In PSO, the cluster fellows cooperate and share data to develop the best solution. By computing each bird's position and speed, a function value calculates the food density of each approaching location. In each search, the direction and speed of the investigation are adapted according to the difference between the best location for its history search and the best location for the population's history search. Ultimately, the entire bird swarm can assemble around the population's ideal site, leading to the discovery of the perfect solution [22]. The processing steps of PSO are listed as follows,

- Random initialization of particle positions and velocity is done in the space of velocity and search.
- The global optimal is generated from the individual optimal solutions for each particle, each of which has a distinct ideal solution, after the fitness function is adjusted. Whether or not the global optimal is updated will depend on the comparison's results. Next, a comparison is made between the current global optimal and the historical global optimal.
- The update of each particle's velocity(v) and position(x) is expressed in Equation 3& 4

$$v_{id} = \omega v_{id} + c_1 r_1 (p_{best,id} - x_{id}) + c_2 r_2 (g_{best,id} - x_{id}) \quad (3)$$

$$x_{id} = x_{id} + v_{id} \quad \text{i=1, 2, 3….Q,} \quad \text{d=1,2….M} \quad (4)$$

where $pbest$ is the individual particle's best value, d shows the dimension of particle i, $gbest$ is the global best value, ω is the inertia factor c_1 and c_2 are the acceleration factor, r_1 and r_2 are the random numbers range from [0, 1].

3. Improved Particle Swarm Optimization algorithm (IPSO)

PSO can considerably decrease the probability of reaching the local optimum prematurely while addressing the optimization issues. We need an enhanced PSO in order to strike the right balance between exploration and exploitation. We suggest an IPSO algorithm with two modifications, "Elimination Mechanism" (EM) and ``Opposition-Based Learning Method (OBL)" to improve the PSO algorithm's performance [23]. To avoid the algorithm from entering the local best value, we first update the swarm group in accordance with the EM principle. After each algorithm iteration, we sort the fitness values corresponding to each swarm in ascending order and delete *R* swarms with the lowest fitness values. Meanwhile, the OBL method generates swarms equal to the number of eliminated swarms.

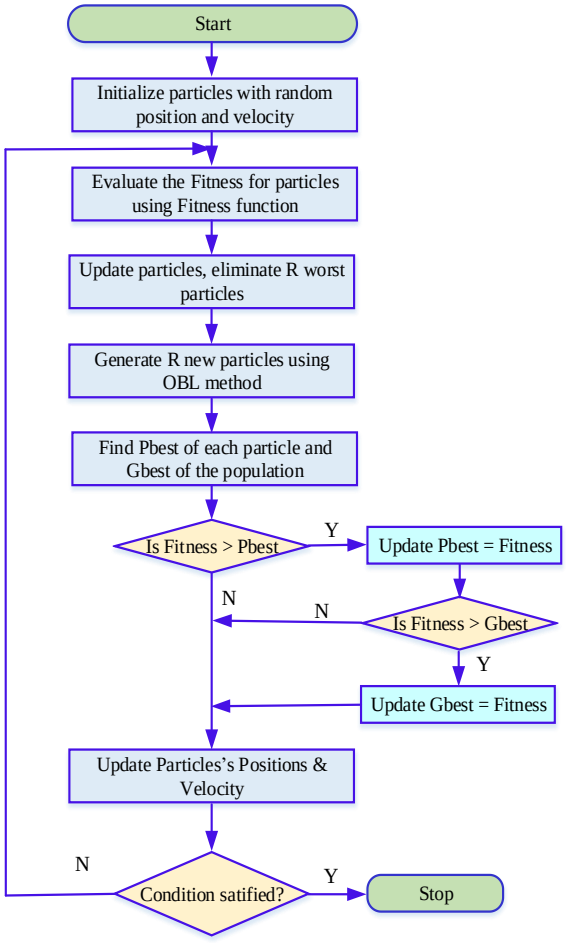


Figure 5. Flowchart of IPSO algorithm

Figure 5 displays the IPSO algorithm's main flowchart. In order to minimize the number of attributes and to enhance detection accuracy, we develop a fitness function. Using equation 5, the fitness function is computed, with a weight factor β falling between 0 and 1.

$$Fitnessfunction = \beta_1 \times Acc + (1-\beta_1) \times \frac{Selectedfeature}{Totalfeature} \quad (5)$$

Improve Particle Swarm Optimization (IPSO)

Input: N: population dimensions

P_l : Local optimal spot

P_g : Group optimal spot

F: fitness function

Output: P_g

1: Randomly set the spot α_i and velocity v_i of particle i
2: **while** criterion is not converge **do**
3: **For** i=1 to N **do**
4: Calculate the fitness value of each swarms according to the fitness function
5: **For** (i=0; i < F_{best} ; i++)
6: **For** (j=0; j<E; j++)
7: Eliminate R worst swarms
8: Generate R new swarms using OBL method
9: **end**
10: **end**
11: Evaluate the fitness function
12: **if** F(α_i) ≥ F(P_l) **then**
13: $P_l \leftarrow \alpha_i$
14: **if** F(α_i) ≥ F(P_g) **then**
15: $P_g \leftarrow P_l$
16: **end**
17: **end**
18: update the position and velocity of particle i
19: Return P_g

4. Feature selection using Chi square-IPSO

Feature selection using Chi-IPSO has been implemented using two stages. The optimum features are ranked according to the chi square method and applied to the input of the IPSO technique. The outcome of the IPSO algorithm is a sequence of 1's and 0's, where the attribute selection is denoted by one and attribute rejection is denoted by zero. The hybrid feature selection principle is shown in Figure 6. The prominent 10 features selected using Chi-IPSO are listed in Table II.

Chi square – IPSO method

Input: Attributes, specification
Output: Best attributes
1: Attributes = Chi square (Attributes)
2: Problem = Attributes
3: Best attributes = IPSO (Problem, specification)
4: Return best attributes

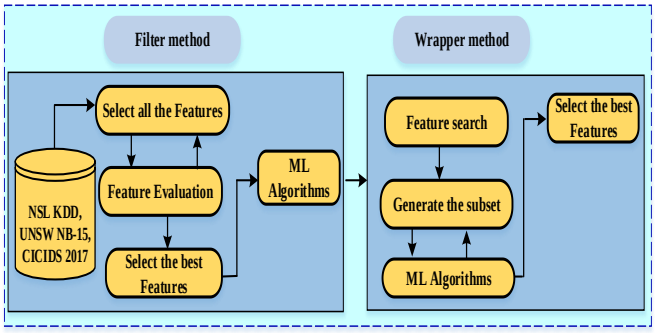


Figure 6. Hybrid feature selection method

TABLE II. LIST OF SELECTED FEATURES BY CHI-IPSO METHOD

S.NO	UNSW NB 15	CICIDS 2017
1	smean	Total Length of Bwd Packets
2	tcprtt	Sub flow Fwd Bytes
3	dttl	Fwd Packet Length Max
4	sbytes	Destination Port
5	ct state ttl	Bwd Header Length
6	sinpkt	Init Win bytes backward
7	res body length	Fwd Packet Length Std
8	ct ftp cmd	Bwd Packet Length Max
9	dbytes	Packet Length Variance
10	din pkt	Init Win bytes forward

D. Ensemble learning algorithm

The ensemble learning algorithm boosts efficiency by breaking down a significant problem into numerous minor and superficial problems that are simpler to comprehend and solve using the divide-and-conquer method. In most situations, the EL algorithm performs better than the ML method. Our proposed system uses the EL algorithms: RF, LGBM, Ada Boosting, and XG Boosting algorithms.

1. Random Forest (RF)

It is an EL algorithm used for efficient attacks and anomaly detection in IoT [24]. Several DT methods are combined to create the RF algorithm, which produces a better result overall. Considering the input dataset having N samples and M features, the dataset is split into K number of new bootstrap dataset. The result obtained

from the individual dataset is combined to produce the concluding result. The working structure is shown in Figure 7.

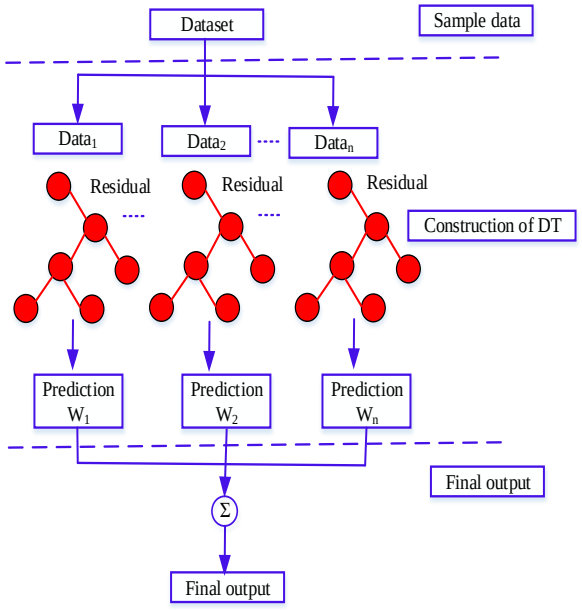


Figure 7. Random forest tree spliting

2.Adaptive boosting

Ada boosting combines several weak learners and produces strong learners. The performance is boosted by iteratively changing the weight of the classifier. The performance will increase depending on the selected base classifier and the voting weight. The algorithm is shown below for the input dataset {(α1,β1), (α2,β2),…. (αn,βn)} [25]

Ada Boosting

Transform weights $We_t(i) = 1/n$; for i= 1, 2, 3, 4, 5…n.

For t=1, 2, 3, 4, 5 …T_1;

{

Using $We_t(i)$ train a DT algorithm $D_t(\alpha)$

Select $D_t(\alpha)$ with low weight error

$$E_t = \sum_{i=1}^{n} We_t(i) I(y_i \neq D_t(\alpha))$$

Specify $\lambda_t = \frac{1}{2}\ln\{(1-E_t)/E_t\}$

Update weight $We_{t+1}(i) = We_t(i)\exp\{-\lambda_t y_i D_t(\alpha_i)/Z_t\}$

// Z_t - Normalization factor

Boosted output as follows

$$H(\alpha) = sign\sum_{t=1}^{Total} \lambda_t D_t(\alpha)$$

}

3. Extreme Gradient Boosting (XG Boosting)

XG Boost is an EL algorithm used in various domain like anomaly detection, attack detection, and feature selection problems. In the XG Boost algorithm, trees are expanded level-wise. The basic blocks of the XG boost algorithms are DT and gradient boosting. From the training data, a prediction is made for the probability of observing normal or anomaly data. The XG Boost algorithm's goal is to determine which cost objective function is optimal. Equation 6 represents the objective function. Where *l* is the loss function, *n* is the number of tree, α is a regularization term, $\hat{y}_i$ is a predicted value, and *s* is the number of samples in the training data [18].

$$\Psi(\theta) = \sum_{i=1}^{s} l(y_i, \hat{y}_i) + \sum_{j=1}^{n} \alpha(f_k) \qquad (6)$$

4.Light Gradient Boosting Machine (LGBM)

It is a leaf-based algorithm to raise trees vertically. The trees are expanded by considering where the loss is minimum during the splitting of tree. Using a histogram-based method, LGBM chooses the best split candidates. The significance of data instances are highlighted by LGBM using the sampling algorithm gradient-based one-side sampling (GOSS) to improve training. Its main objective is to ignore minor gradients and concentrate on data samples with massive ones. Since data with short gradients have less errors, the underlying assumption is that they have already been trained. GOSS recommended eliminating these uninformative samples and using the remaining samples to calculate the knowledge obtained while finding the appropriate divides. Unfortunately, this will cause a bias issue in favor of the sample with more giant gradients and alter the initial data distribution. To address this issue, GOSS keeps all the samples with big gradients while randomly selects the data with slight gradients. For the purpose of computing the information gain, GOSS increases the weights of the data instances with small gradients [26].

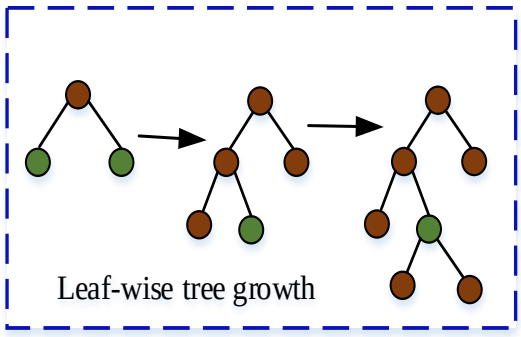


Figure 8. LGBM tree growth

E. *Assessment parameters*

The performance of the developed method is assessed using common ML algorithm evaluation metrics extrapolated from the confusion matrix. Figure 9 shows the confusion matrix of our classification problem, where D11 denotes the number of anomaly that are correctly estimated, D12 denotes the number of anomaly incorrectly predicted, D21 and D22 denotes the number of non-anomaly that were estimated incorrectly and correctly respectively. Equation (7-11) lists the evaluation parameters for our proposed model.

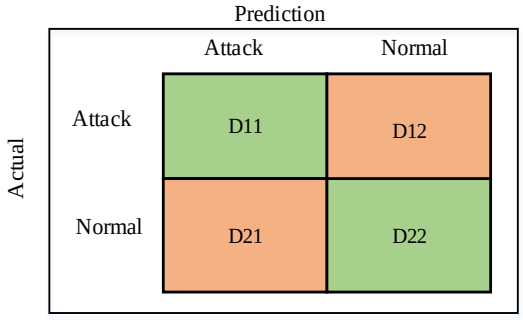


Figure 9. Confusion matrix

$$Accuracy = \frac{D11 + D22}{D11 + D22 + D12 + D21} \quad (7)$$

$$precision = \frac{D11}{D11 + D21} \quad (8)$$

$$\text{Senstivity} = \frac{D11}{D11 + D12} \quad (9)$$

$$F1score = \frac{2 * (P * S)}{(P + S)} \quad (10)$$

$$MCC = \frac{D11 * D22 - D12 * D21}{\sqrt{(D11 + D21)(D11 + D12)(D22 + D21)(D22 + D12)}} \quad (11)$$

4. RESULT ANALYSIS

A. *Experimental setup*

This research uses a Dell system with Linux OS 64-bit. The PC has the processor of Intel(R) Core i5 with 8GB RAM. The attribute selection and ML algorithms are executed using Python with a Co-lab environment. Scikit Learn and Pyswarm library are used in our experimental analysis. The developed model was trained and tested using 20-fold CV method. The total of 24 experimental combinations are created with the use of four EL algorithms (RF, Ada Boost, LGBM, and XG Boost) and two intrusion datasets (UNSW NB 15, and CICIDS 2017).

In [27], Tama et al. performed an improved ADS in web traffic using EL algorithms. The author proved that the performance of ADS is improved by proper parameter-tuning of EL algorithms. PSO algorithm is used to detect the optimum hyper parameters of EL algorithms. The initial settings of the PSO algorithm are shown in Table III.

TABLE III. INITIAL SETTING OF PSO METHOD

S.NO	Parameters	Values
1	C_1, C_2	2
2	Number of populations	30
3	Inertia factor	0.7
4	Number of iterations	100

B. *Result discussion*

We have used four EL algorithms (RF, Ada Boost, LGBM, and XG Boost) with default parameters to compare the performance in this anomaly detection analysis. Tables IV and V deliver the experimental assessment outcomes of the four EL-based anomaly detection results for various datasets with full features. The RF algorithm's performance is better than the remaining algorithms in terms of test accuracy, precision, sensitivity, F-score, and MCC value for all dataset.

Despite the increase in prediction accuracy, it is necessary to increase the detection accuracy of the EL algorithm. Hyper-parameter tuning is a crucial step to improve detection accuracy and save training time. The popular PSO algorithm enhances anomaly detection performance by fine-tuning the EL algorithm's parameters. Using the PSO method, we carefully tune several EL algorithm parameters, including tree size, learning rate, and maximum depth, which are tabulated in Tables VI and VII. Comparing the Tables IV and V and Tables VI and VII, the performance of the tuned EL algorithm is better, which is considered for further model design.

The developed model has been used to conduct numerous experiments by tuning the number of features for each training session. Initially, all the features are ranked rapidly by applying the chi-square method. The top 16 most relevant features are selected and applied as the input of the developed IPSO algorithm. After several iterations, we select the ten most prominent features for model building. Our proposed model selects the critical features by combining the advantages of filtering (Chi-square) and wrapper method (IPSO). Tables VIII-IX show the detection results of the proposed method (Chi-IPSO-based feature selection) with optimized ensemble learning algorithm on two datasets.

TABLE IV. THE DETECTION RESULTS OF FULL FEATURES WITH DEFAULT PARAMETER EL ALGORITHM ON THE UNSW NB 15

EL algorithm	Accuracy	Precision	Sensitivity	MCC	F1-score
ADA BOOST	0.9268	0.9294	0.9590	0.6002	0.9588
LGBM	0.9378	0.9351	0.9711	0.6017	0.9651
XG BOOST	0.9399	0.9365	0.9768	0.6050	0.9677
RF	0.9402	0.9371	0.9769	0.6075	0.9697

TABLE V. THE DETECTION RESULTS OF FULL FEATURES WITH DEFAULT PARAMETER EL ALGORITHM ON THE CICIDS 2017

EL algorithm	Accuracy	Precision	Sensitivity	MCC	F1-score
ADA BOOST	0.9790	0.9764	0.9810	0.9581	0.9787
LGBM	0.9799	0.9804	0.9788	0.9599	0.9796
XG BOOST	0.9891	0.9886	0.9891	0.9781	0.9889
RF	0.9905	0.9907	0.9901	0.9811	0.9904

TABLE VI. THE DETECTION RESULTS OF FULL FEATURES WITH TUNED EL ALGORITHM ON THE UNSW NB 15

EL algorithm	Accuracy	Precision	Sensitivity	MCC	F1-score
ADA BOOST	0.9366	0.9316	0.9653	0.6013	0.9637
LGBM	0.9401	0.9474	0.9724	0.6108	0.9709
XG BOOST	0.9432	0.9481	0.9732	0.6113	0.9724
RF	0.9434	0.9498	0.9812	0.6202	0.9817

TABLE VII. THE DETECTION RESULTS OF FULL FEATURES WITH TUNED EL ALGORITHM ON THE CICIDS 2017

EL algorithm	Accuracy	Precision	Sensitivity	MCC	F1-score
ADA BOOST	0.9802	0.9721	0.9853	0.9603	0.9886
LGBM	0.9880	0.9901	0.9912	0.9761	0.9922

XG BOOST	0.9926	0.9925	0.9902	0.9852	0.9930
RF	0.9955	0.9955	0.9946	0.9911	0.9950

TABLE VIII. THE DETECTION RESULTS OF OUR PROPOSED METHOD ON THE UNSW NB 15

EL algorithm	Accuracy	Precision	Sensitivity	MCC	F1-score
ADA BOOST	0.9394	0.9394	0.9764	0.6163	0.9642
LGBM	0.9441	0.9517	0.9752	0.6250	0.9763
XG BOOST	0.9454	0.9574	0.9780	0.6275	0.9802
RF	0.9458	0.9590	0.9855	0.6328	0.9859

TABLE IX. THE DETECTION RESULTS OF OUR PROPOSED METHOD ON THE CICIDS 2017

EL algorithm	Accuracy	Precision	Sensitivity	MCC	F1-score
ADA BOOST	0.9846	0.9812	0.9832	0.9692	0.9856
LGBM	0.9903	0.9932	0.9900	0.9806	0.9906
XG BOOST	0.9940	0.9945	0.9912	0.9881	0.9934
RF	0.9970	0.9964	0.9932	0.9940	0.9955

The performances of the UNSW NB 15 dataset with full features using Ada boost, LGBM, XG Boost, and RF classifiers are 92.68%, 93.78%, 93.99%, and 94.02%, respectively. Similarly, the performances achieved with the feature selection technique using Ada boost, LGBM, XG Boost, and RF classifiers are 93.94%, 94.41%, 94.54%, and 94.58%, respectively. The performances of the CICIDS 2017 dataset without using feature selection techniques using Ada boost, LGBM, XG Boost, and RF classifiers are 97.90%, 97.99%, 98.91%, and 99.05%, respectively. Similarly, the performances achieved with the feature selection technique using Ada boost, LGBM, XG Boost, and RF classifiers are 99.05% to 98.46%, 99.03%, and 99.40%, respectively. The above discussion infers that the performance of the classifiers with selected attributes is superior to the performance of the classifiers without FS technique, irrespective of the dataset. Figure 11 represents the ROC curve of Ada Boost, LGBM, XG Boost, and RF algorithm for two datasets. From the figure, it can be observed that RF has a higher AUC value for all the three dataset.

Figure 10 displays the performance comparison of the developed model for all different datasets (UNSW NB 15, CICIDS 2017) using various evaluation metrics such as accuracy, precision, recall, and F-score. From the figure, the RF algorithm has better performance compared to the Ada boost, LGBM, and XG Boost algorithms for all datasets.

Figure 11 (a) displays the ROC curve of the UNSW NB 15 dataset for several EL algorithms. The respective AUC values for each classifier are Ada boost (0.97), LGBM (0.97), XG boost (0.97), and RF classifier (0.98). It is evident that the RF algorithm has the superior AUC value among these classifiers. Figure 11 (b) shows the ROC curve of various EL algorithms for the CICIDS 2017 dataset. The AUC values for the individual classifiers are as follows: Ada boost-0.99, LGBM-0.99, XG Boost-1 and RF classifier-1. Among these classifiers, the XG Boost and RF algorithms have the best AUC values. Figure 11 (c) shows the consolidated ROC curve of different EL algorithms on two different datasets. The AUC values of the different classifiers are denoted individually in the figure for further comparison.

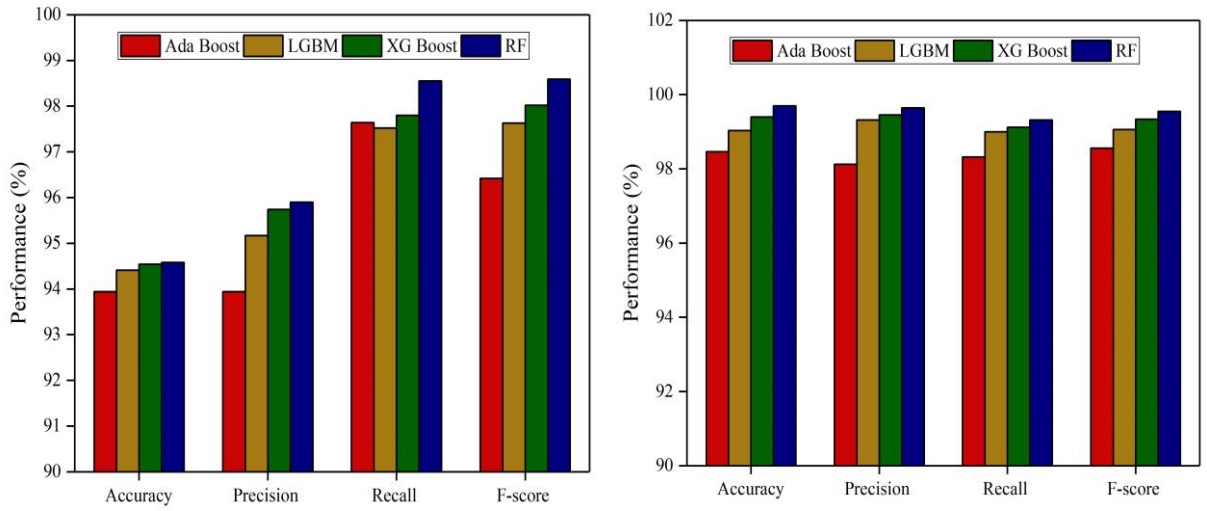
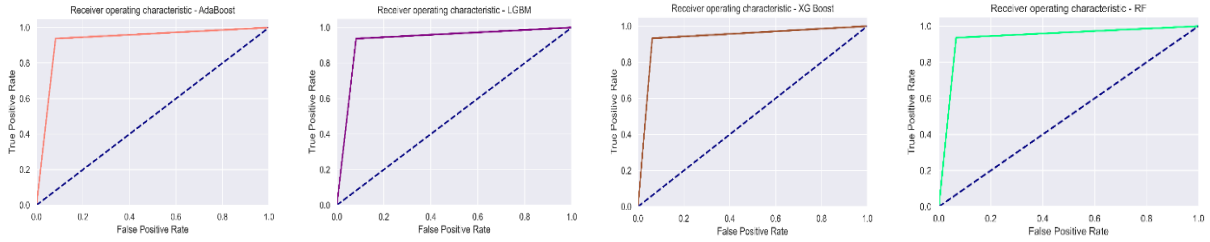
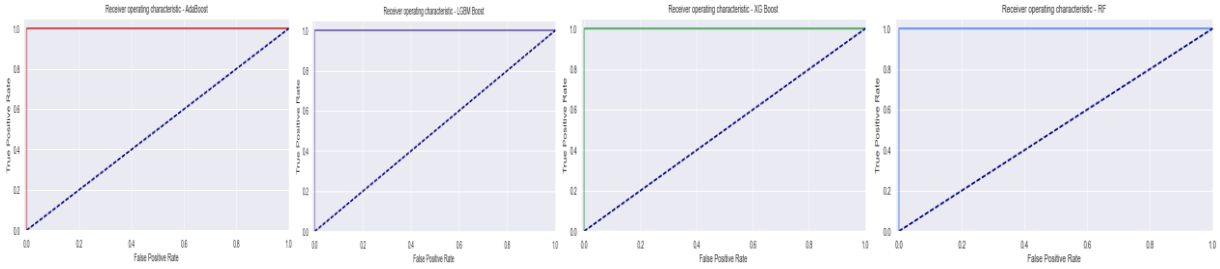


Figure 10. Performance comparison of proposed model: (a) UNSW NB 15 (b) CICIDS 2017



(a)



(b)

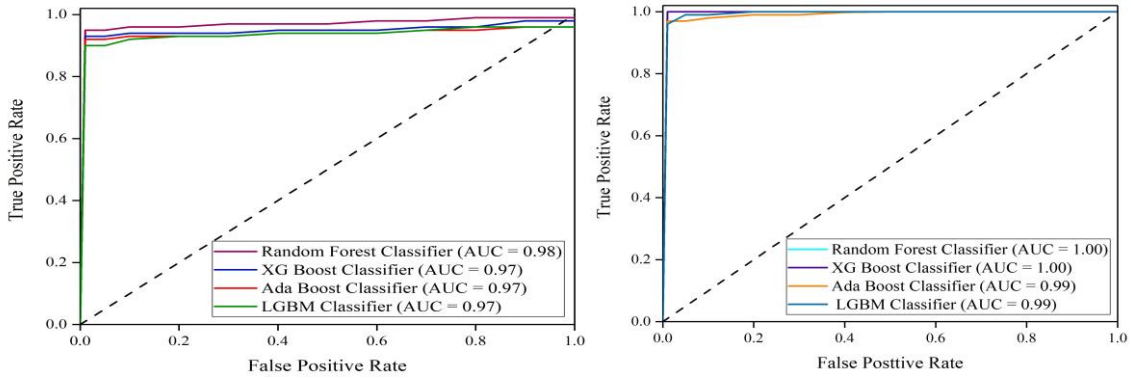


Figure 11. ROC Curve of different EL algorithm on different dataset: (a) UNSW NB-15 (b) CICIDS 2017 (c) All the Classifiers

The MCC value comparison of the proposed Chi-IPSO feature selection technique with EL algorithms for different datasets is shown in Figure 12. For the UNSW NB 15 dataset, the MCC values for the different classifiers are: ada boost (0.6163), LGBM (0.6250), XG Boost (0.6275), and RF classifier (0.6328). For the CICIDS 2017 dataset, the MCC values for the different classifiers are: ada boost-0.9692, LGBM-0.9806, XG boost -0.9881, and RF classifier-0.9940. From the figure, the RF algorithm has the highest MCC values of 0.6328, and 0.9949 for the UNSW NB 15, and CICIDS 2017 datasets. This means the RF algorithm can forecast the result more accurately than other algorithms.

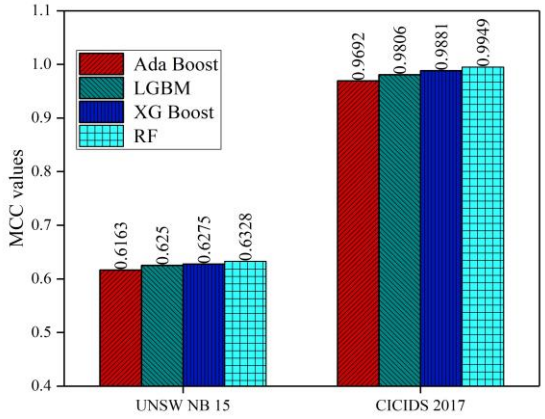


Figure 12. MCC value comparison of different EL algorithms

C. Comparative study

When the developed model is compared with the contemporary ADS, the proposed model can be identified with various strengths and weaknesses, which enable further improvements in our model. B A Tama et al. [28] created an IDS using a hybrid ensemble learning algorithm and achieved 90.39% accuracy value for the NSL KDD dataset using PSO + Gradient boosting algorithm. Joydip et al. [29] suggested an IDS using PSO with various ML algorithms, among them RF + PSO produced an excellent detection rate of 99.26%. Qusyairi et al. [9] developed an anomaly-based ADS using Spearman's rank correlation coefficient feature selection with an ensemble learning algorithm. For CIC-IDS 2018 dataset, this system achieved 98.8% accuracy and 97.9 % F1 score. Orieb Abu Alghanam et al. [14] developed an IDS in IoT using a novel improved PIO feature selection techniques with iForest + SVM classifier and achieved 96.93% accuracy for the NSL KDD dataset. Kumar et al. [30] offered an efficient IDS using two new optimization techniques for FS, namely the binary gravitational search algorithm and GWO algorithm, and attained an accuracy of 99.41%. Additionally, we evaluated the developed Chi-IPSO-RF method by comparing it with other existing models and verified it on the benchmark dataset. Table 15 reveals that our proposed model is superior to most of the existing IDS at IoT.

TABLE X. COMPARISON TABLE OF THE DEVELOPED MODEL WITH PRESENT IDS MODELS.

Study	Dataset	Model	Accuracy
[11]	NSL KDD	RF, DT	86.68%
[12]	CICIDS 2017	Voting(RF, SVM, NB, KNN)	99.70%
[17]	CIC-IDS 2017, NSL KDD, UNSW NB 15	Voting (DT, NB, LR, NN, SVM)	99.50%, 88.10%, 85.70%
[31]	UNSW NB 15	SMO-HPSO	94.12%
[32]	UNSW NB 15	LGBM	85.89%
[33]	NSL KDD	B-Stacking	98.50%
[34]	UNSW NB 15	IGRF-RFE	84.24%
[35]	CICIDS 2017	CNN-GRU	98.73%
[36]	CICIDS 2017	XG Boost	98.00%
[37]	CICIDS 2017	CI-EnsID	97.90%

Study	Dataset	Model	Accuracy
[38]	UNSW NB 15, CICIDS 2017	Stacking	93.88%, 99.80%
[39]	UNSW NB15, CICIDS 2017	DNN	96.70%, 98.74%
[40]	NSL KDD	Voting (SVM, LR, NB, DT)	96.06%
[41]	UNSW NB 15	Voting (SVM, DT, ANFIS)	98.34%
Proposed	UNSW NB15, CICIDS 2017	Chi-IPSO-RF	94.58%, 99.70%,

5. CONCLUSIONS AND FUTURE WORKS

In this paper, we propose an effective anomaly detection system in IoT using a Chi-IPSO feature selection with an optimum ensemble learning algorithms. The newly developed Chi-IPSO algorithm selects dominant features from the IoT dataset. To address the shortcomings of basic PSO, which easily falls into the local optimum and has a slow convergence speed, we propose an IPSO algorithm, which is then hybridized with the chi-square method. The proposed Chi-IPSO feature selection technique reduces the number of features in the UNSW NB 15 dataset from 49 to 10. Similarly, it reduces the number of features from 79 to 10 for the CICIDS 2017 dataset. Furthermore, we utilized various ensemble classifiers to examine classification errors. Finally, the simulation results reveal that the RF classifier with the proposed feature selection technique achieves the highest accuracy of 94.58% for the UNSW NB 15, and 99.70% for the CICIDS 2017 dataset. In the future, this study can be expanded using different feature selection techniques, novel balancing techniques, and deploying the model in real time to categorize network data.

ACKNOWLEDGMENT

The authors thank Natural Sciences and Engineering Research Council of Canada (NSERC) and New Brunswick Innovation Foundation (NBIF) for the financial support of the global project. These granting agencies did not contribute in the design of the study and collection, analysis, and interpretation of data.

REFERENCES

[1] G Kumar, Sachin, Prayag Tiwari, and Mikhail Zymbler. “Internet of Things is a revolutionary approach for future technology enhancement: a review." Journal of Big data 6, no. 1 (2019): 1-21, https://doi.org/10.1186/s40537-019-0268-2.

[2] E Hussain, Md Iftekhar. "Internet of Things: challenges and research opportunities." CSI transactions on ICT 5 (2017): 87-95, 10.1007/s40012-016-0136-6.

[3] Hassija, Vikas, Vinay Chamola, Vikas Saxena, Divyansh Jain, Pranav Goyal, and Biplab Sikdar. “A survey on IoT security: application areas, security threats, and solution architectures” IEEE Access 7 (2019): 82721-82743, , https://doi.org/10.1109/access.2019.2924045 .

[4] Manokaran, J., and G. Vairavel. “Smart anomaly detection using data-driven techniques in iot edge: a survey. ” In Proceedings of Third International Conference on Communication, Computing and Electronics Systems, pp. 685-702. Singapore: Springer Singapore, 2022, 10.1007/978-981-16-8862-1_45.

[5] M, J., and G. V. "An empirical comparison of machine learning algorithms for attack detection in internet of things edge." *ECS Transactions* 107, no. 1 (2022): 2403 , https://doi.org/10.1149/10701.2403ecst.

[6] Manokaran, J., and G Vairavel. "GIWRF-SMOTE: Gini impurity-based weighted random forest with SMOTE for effective malware attack and anomaly detection in IoT-Edge." *Smart Science* 11, no. 2 (2023): 276-292, DOI: 10.1080/23080477.2022.2152933.

[7] Tama, Bayu Adhi, and Sunghoon Lim. "Ensemble learning for intrusion detection systems: A systematic mapping study and cross-benchmark evaluation." *Computer Science Review* 39 (2021): 100357, https://doi.org/10.1016/j.cosrev.2020.100357.

[8] Maldonado, Javier, María Cristina Riff, and Bertrand Neveu. "A review of recent approaches on wrapper feature selection for intrusion detection." *Expert Systems with Applications* 198 (2022): 116822, https://doi.org/10.1016/j.eswa.2022.116822.

[9] Fitni, Qusyairi Ridho Saeful, and Kalamullah Ramli. "Implementation of ensemble learning and feature selection for performance improvements in anomaly-based intrusion detection systems." In *2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, pp. 118-124. IEEE, 2020, doi: 10.1109/IAICT50021.2020.9172014.

[10] Albulayhi, Khalid, Qasem Abu Al-Haija, Suliman A. Alsuhibany, Ananth A. Jillepalli, Mohammad Ashrafuzzaman, and Frederick T. Sheldon. "IoT intrusion detection using machine learning with a novel high performing feature selection method." *Applied Sciences* 12, no. 10 (2022): 5015, https://doi.org/10.3390/app12105015.

[11] Abdullah, Manal, Arwa Alshannaq, Asmaa Balamash, and Soad Almabdy. "Enhanced intrusion detection system using feature selection method and ensemble learning algorithms." *International Journal of Computer Science and Information Security (IJCSIS)* 16, no. 2 (2018): 48-55.

[12] Mhawi, Doaa N., Ammar Aldallal, and Soukeana Hassan. "Advanced feature-selection-based hybrid ensemble learning algorithms for network intrusion detection systems." *Symmetry* 14, no. 7 (2022): 1461, https://doi.org/10.3390/sym14071461.

[13] Rahman, Md Arafatur, A. Taufiq Asyhari, Ong Wei Wen, Husnul Ajra, Yussuf Ahmed, and Farhat Anwar. "Effective combining of feature selection techniques for machine learning-enabled IoT intrusion detection." *Multimedia Tools and Applications* (2021): 1-19, https://doi.org/10.1007/s11042-021-10567-y.

[14] Alghanam, O A, Wesam Almobaideen, Maha Saadeh, and Omar Adwan. "An improved PIO feature selection algorithm for IoT network intrusion detection system based on ensemble learning." *Expert Systems with Applications* 213 (2023): 118745, https://doi.org/10.1016/j.eswa.2022.118745.

[15] Kaushik, Baijnath, Reya Sharma, Kulwant Dhama, Akshma Chadha, and Surbhi Sharma. "Performance evaluation of learning models for intrusion detection system using feature selection." *Journal of Computer Virology and Hacking Techniques* (2023): 1-20, https://doi.org/10.1007/s11416-022-00460-z.

[16] Jaw, Ebrima, and Xueming Wang. "Feature selection and ensemble-based intrusion detection system: an efficient and comprehensive approach." *Symmetry* 13, no. 10 (2021): 1764, https://doi.org/10.3390/sym13101764.

[17] Das, Saikat, Sajal Saha, Annita Tahsin Priyoti, Etee Kawna Roy, Frederick T. Sheldon, Anwar Haque, and Sajjan Shiva. "Network intrusion detection and comparative analysis using ensemble machine learning and feature selection." *IEEE Transactions on Network and Service Management* (2021), doi: 10.1109/TNSM.2021.3138457.

[18] Jiang, Hui, Zheng He, Gang Ye, and Huyin Zhang. "Network intrusion detection based on PSO-XGBoost model." *IEEE Access* 8 (2020): 58392-58401, doi: 10.1109/ACCESS.2020.2982418.

[19] Moustafa, Nour, and Jill Slay. "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)." In *2015 military communications and information systems conference (MilCIS)*, pp. 1-6. IEEE, 2015, doi: 10.1109/MilCIS.2015.7348942.

[20] Sharafaldin, Iman, Arash Habibi Lashkari, and Ali A. Ghorbani. "Toward generating a new intrusion detection dataset and intrusion traffic characterization." *ICISSp* 1 (2018): 108-116, https://www.scitepress.org/papers/2018/66398/66398.pdf.

[21] Thaseen, I. Sumaiya, Ch Aswani Kumar, and Amir Ahmad. "Integrated intrusion detection model using chi-square feature selection and ensemble of classifiers." *Arabian Journal for Science and Engineering* 44 (2019): 3357-3368, https://doi.org/10.1007/s13369-018-3507-5.

[22] Louk, Maya Hilda Lestari, and Bayu Adhi Tama. "PSO-Driven Feature Selection and Hybrid Ensemble for Network Anomaly Detection." *Big Data and Cognitive Computing* 6, no. 4 (2022): 137, https://doi.org/10.3390/bdcc6040137 .

[23] Manokaran, J., and G. Vairavel. "IGWO-SoE: Improved Grey Wolf Optimization based Stack of Ensemble Learning Algorithm for Anomaly Detection in Internet of Things Edge Computing." *IEEE Access* (2023), doi.org/10.1109/access.2023.3319814 .

[24] Manokaran, J., G. Vairavel, and J. Vijaya. "A Novel Set Theory Rule based Hybrid Feature Selection Techniques for Efficient Anomaly Detection System in IoT Edge." In *2023 International Conference on Quantum Technologies, Communications, Computing, Hardware and Embedded Systems Security (iQ-CCHESS)*, pp. 1-6. IEEE, 2023.

[25] Sharma, Manoj. "Cervical cancer prognosis using genetic algorithm and adaptive boosting approach." *Health and Technology* 9 (2019): 877-886, https://doi.org/10.1007/s12553-019-00375-8 .

[26] Alzamzami, Fatimah, Mohamad Hoda, and Abdulmotaleb El Saddik. "Light gradient boosting machine for general sentiment classification on short texts: a comparative evaluation." *IEEE access* 8 (2020): 101840-101858, , doi: 10.1109/ACCESS.2020.2997330.

[27] Tama, Bayu Adhi, Lewis Nkenyereye, SM Riazul Islam, and Kyung-Sup Kwak. "An enhanced anomaly detection in web traffic using a stack of classifier ensemble." *IEEE Access* 8 (2020): 24120-24134, doi: 10.1109/ACCESS.2020.2969428.

[28] Louk, Maya Hilda Lestari, and Bayu Adhi Tama. "PSO-Driven Feature Selection and Hybrid Ensemble for Network Anomaly Detection." *Big Data and Cognitive Computing* 6, no. 4 (2022): 137.

[29] Kunhare, Nilesh, Ritu Tiwari, and Joydip Dhar. "Particle swarm optimization and feature selection for intrusion detection system." *Sādhanā* 45 (2020): 1-14, https://doi.org/10.1007/s12046-020-1308-5.

[30] Dey, Arun Kumar, Govind P. Gupta, and Satya Prakash Sahu. "A metaheuristic-based ensemble feature selection framework for cyber threat detection in IoT-enabled networks." *Decision Analytics Journal* 7 (2023): 100206 https://doi.org/10.1016/j.dajour.2023.100206.

[31] Turukmane, Anil V., and Ramkumar Devendiran. "M-MultiSVM: An efficient feature selection assisted network intrusion detection system using machine learning." *Computers & Security* 137 (2024): 103587, https://doi.org/10.1016/j.cose.2023.103587.

[32] Liu, Jingmei, Yuanbo Gao, and Fengjie Hu. "A fast network intrusion detection system using adaptive synthetic oversampling and LightGBM." *Computers & Security* 106 (2021): 102289, https://doi.org/10.1016/j.cose.2021.102289.

[33] Roy, Souradip, Juan Li, Bong-Jin Choi, and Yan Bai. "A lightweight supervised intrusion detection mechanism for IoT networks." *Future Generation Computer Systems* 127 (2022): 276-285, https://doi.org/10.1016/j.future.2021.09.027.

[34] Yin, Yuhua, Julian Jang-Jaccard, Wen Xu, Amardeep Singh, Jinting Zhu, Fariza Sabrina, and Jin Kwak. "IGRF-RFE: a hybrid feature selection method for MLP-based network intrusion detection on UNSW-NB15 dataset." *Journal of Big Data* 10, no. 1 (2023): 1-26, https://doi.org/10.1186/s40537-023-00694-8.

[35] Henry, Azriel, Sunil Gautam, Samrat Khanna, Khaled Rabie, Thokozani Shongwe, Pronaya Bhattacharya, Bhisham Sharma, and Subrata Chowdhury. "Composition of hybrid deep learning model and feature optimization for intrusion detection system." *Sensors* 23, no. 2 (2023): 890.

[36] OYELAKIN, Akinyemi Moruff. "A Learning Approach for The Identification of Network Intrusions Based on Ensemble XGBoost Classifier." *Indonesian Journal of Data and Science* 4, no. 3 (2023): 190-197.

[37] Muhammad, Ali, Iqbal Murtza, Ayesha Saadia, and Kashif Kifayat. "Cortex-inspired ensemble based network intrusion detection system." *Neural Computing and Applications* (2023): 1-14, https://doi.org/10.1007/s00521-023-08561-6.

[38] Thockchom, Ngamba, Moirangthem Marjit Singh, and Utpal Nandi. "A novel ensemble learning-based model for network intrusion detection." *Complex & Intelligent Systems* (2023): 1-22.

[39] Thakkar, Ankit, and Ritika Lohiya. "Attack classification of imbalanced intrusion data for IoT network using ensemble learning-based deep neural network." *IEEE Internet of Things Journal* (2023), doi: 10.1109/JIOT.2023.3244810.

[40] Krishnaveni, Sivamohan, Sivanandam Sivamohan, S. S. Sridhar, and S. Prabakaran. "Efficient feature selection and classification through ensemble method for network intrusion detection on cloud computing." *Cluster Computing* 24, no. 3 (2021): 1761-1779, https://doi.org/10.1007/s10586-020-03222-y.

[41] Vanitha, S., and P. Balasubramanie. "Improved Ant Colony Optimization and Machine Learning Based Ensemble Intrusion Detection Model." *Intelligent Automation & Soft Computing* 36, no. 1 (2023), doi:10.32604/iasc.2023.032324.