Journal of Al-Qadisiyah for Computer Science and Mathematics Vol. 16(3) 2024, pp Comp. 84–95

Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

A Novel Hybrid Intrusion Detection Framework Leveraging Machine Learning and Flower Pollination Optimization

Shahad Ali Sameer[a], Hadeel Qassim Albaaj[a], Safa Jaber Abbas[a,*], Maryam Najeh Khalill[a], and Areej Qassim[a]

College of Computer Science and Information Technology . Al-Qadisah University , Diwaniyah-Iraq .Email: Shahadalisameer@gmail.com, hadelqassim95@gmail.com, safa.abo.tabikh@qu.edu.iq, maryamnajeh93@gmail.com, Enggoga6@gmail.com

ARTICLEINFO

Article history:
Received: 18 /07/2024
Rrevised form: 31 /07/2024
Accepted : 26 /08/2024
Available online: 30 /09/2024

Keywords: Intrusion detection system, Machine learning, Feature selection, Flower Pollination Optimization (FPO) .

ABSTRACT

The exponential growth in technologies such as cloud computing, smart devices, virtualization, and the Internet of Things (IoT) has generated over four hundred zettabytes of network traffic data annually. This surge necessitates robust cybersecurity strategies to protect information from intrusions, which can result in significant financial losses. Reducing security risks requires the use of data analytics and machine learning to derive insights and make informed decisions based on network data. This study introduces the Flower Pollination Optimization (FPO) algorithm for feature selection to enhance the performance of several classifiers on the UNSW-NB15 dataset. We evaluated four classifiers: Linear Discriminant Analysis (LDA), Multi-Layer Perceptron (MLP), Quadratic Discriminant Analysis (QDA), and K-Nearest Neighbors (KNN) in two scenarios: without feature selection and with FPO-based feature selection. The results demonstrate significant improvements in classifier performance with FPO, with QDA achieving the highest accuracy of 99.16%. Comparative analysis with recent studies highlights the superior performance of our approach, setting a new benchmark in intrusion detection. This research underscores the essential role of effective feature selection in improving the accuracy and reliability of Intrusion Detection Systems (IDS), particularly in IoT environments.

MSC.

https://doi.org/10.29304/jqcsm.2024.16.31645

1.Introduction

This exponential increase in cyber traffic and unprecedented growth in very sophisticated intrusion attacks require adaptive and timely intrusion detection. Cyber-attacks can manifest themselves by way of unauthorized, phishing, and malware infection-sensitive data access, amongst others, which are all capable of such devastating effects on the organization that would translate into reputational damage, financial loss, or the exposure of sensitive information [1] [2]. Vigilance and proactive steps to protect industries, businesses, and individuals from such cyber threats are of foremost necessity. The adaptive real-time intrusion detection system shall continuously monitor and detect emergent threats in a bid to enable organizations to proactively defend against cyberattacks, thereby minimizing the

∗Corresponding author Safa Jaber Abbas

Email addresses: safa.abo.tabikh@qu.edu.iq

Communicated by 'sub etitor'

impact of breaches [3] [4]. These current methods of intrusion detection cannot support modern cyber threats because of their dynamic and complex nature, with a view to improving accuracy and reducing false positives. Machine learning allows algorithms to be trained on datasets to identify network traffic anomalies [5].

During the Internet of Things (IoT), network traffic has increased, making this beneficial. Data analytics has solidified its place in cybersecurity research, paving the path for the collection, storage, and processing of cybersecurity information other than log and alert analysis from firewalls and intrusion detection systems (IDS). This includes communication data with voice, email and social networking activity, along with files and configuration information of assets integrated for advanced analytics aiding user identification behavior; business process data could also be integrated with the configuration information of assets for risk assessments [6, 7]. Due to significant correlations between data and an imbalanced dataset, these systems produce a lot of false positives, with abnormal data cases outweighing normal ones. The following are the research's primary contributions:

- The Flower Pollination Optimization (FPO) algorithm was introduced for feature selection in the UNSW-NB15 dataset, demonstrating significant improvements in the accuracy and reliability of various classifiers. This approach effectively reduces the dimensionality of the data while maintaining relevant features, leading to better performance in intrusion detection.

- A thorough evaluation of multiple classifiers, including Linear Discriminant Analysis (LDA), Multi-Layer Perceptron (MLP), Quadratic Discriminant Analysis (QDA), and K-Nearest Neighbors (KNN), was conducted under two scenarios: with and without feature selection. The results highlight the superiority of using FPO for feature selection, particularly in enhancing the performance of LDA and QDA classifiers. Notably, QDA achieved the highest accuracy of 99.16%, and LDA achieved an accuracy of 99.02%.

- This work provides a detailed comparison with recent studies on the UNSW-NB15 dataset, showcasing the effectiveness of the proposed method. The comparative analysis underscores that this approach sets a new benchmark in the field, achieving higher accuracy and reliability in detecting anomalies and intrusions in IoT environments.

2. Related Works

Conventional defense systems, such as IDSs and firewalls, need updates on a continuous basis for threat detection, according to [8]. The author advocates that IDS must be equipped with the best machine learning models so that reliability and detection rates go up, making the number of false alarms lower. The paper has handled the problem of imbalanced classes in it through the Synthetic Minority Oversampling Technique to generate an accuracy as high as 99%. In [9], a comparative analysis of classification algorithms for the UNSWNB15 dataset has been conducted. This paper concludes that among all of the classification models, random forest classification seems to be more reliable, with an accuracy as high as 97.49%. In [10], different machine learning models were developed under Apache Spark, running the UNSW NB15 dataset for comparison. The outcome of this research states that the random forest model performs most accurately as compared to the NB and DT models.

In [11], a filter-based feature reduction method has been presented with analysis on the UNSW-NB15 intrusion detection dataset using the XGBoost algorithm. In the present study, various machine learning techniques have been applied to the reduced feature set, namely: decision trees, logistic regression, support vector machines knn, and neural networks. Using XGBoost-based feature selection, the DT-based binary classification scheme's test accuracy increases from 88.13% to 90.85%. According to the authors of [12], network intrusion detection systems are essential for solving internet security problems in an IoT environment. Their study claimed that UNSW-NB15 is a very apt dataset for evaluating the performance of different NIDS. The experimental results show that the SVM method has the best performance, while it achieves an accuracy of 75.77% in multi-classification and 85.99% in binary classification. In the paper [13], some authors present the issue of network intrusion classification, stating that traditional machine learning algorithms become ineffective due to the large volume of feature space. This work promotes deep learning as a better approach for solving this problem because of its ability to handle high dimensions and complex features. In this paper, a feed-forward neural network is proposed that provides 99% accuracy with a reduced false alert rate.

In this type of experiment about network intrusion detection, described in [14], potential cyber threats are being identified. The work analyzes classical, hybrid, and ensemble approaches and establishes a model that stacks machine learning algorithms together with extra tree classifier and mutual information gain feature selection techniques. The model has 96.24% UNSW-NB15 dataset accuracy. [15] proposes a hybrid model by combining the

Random Forest as a feature selection model with a Classification and Regression Trees (CART) classification model. With this model, the important features were identified with an accuracy of 95.37% for classifying different types of attacks. Finally, [16] compares UNSW-NB15 and CICIDS2017. The study found that the Naïve Bayes (NB) feature embedding SVM model achieves accuracy of 98.92% on the UNSW-NB15 dataset.

More concretely, [17] states that legacy defense mechanisms, such as firewalls and IDSs, are assured of updated protection to recognize threats continually. The same work recommends that for IDSs to adopt machine learning models, the accuracy results are to achieve reliability and increase the detection ability by reducing the false alarm. This paper tries to take care of the imbalanced class issue, and the accuracy reached is 99% upon processing the data with the synthetic minority oversampling technique. Spatial and hierarchical features in a dataset are analyzed using convolutional neural networks (CNNs). The long-term temporal features of the data are explored using Bi-LSTM layers. Combining these two methods makes it feasible to anticipate prospective assaults. Stratified k-fold cross validation is used to assess the binary classification outcomes of the suggested model for the UNSW-NB15 dataset. K values may range from 2 to 10. These results include classifying the accuracy at 93.84%, the average detection rate at 94.70%, and the false positive rate at 7.70%, according to the research by Sinha et al. [18]. This means that the proposed approach of deep learning can enhance IDS performance. However, it is delineated that researchers with increased accuracy and robustness, like Jose and Jose [19], have tested LSTM-based IDSs on various datasets. The accuracy of IDS in UNSWNB15 was at most 98.92%. Further improvement in reducing false detection probability is required, since it reduces the accuracy of the total system. Table 1 summarizes recent studies in IDS detection.

Table 1 – Summary of recent research in IDS Detection.

Ref	Year	Preprocessing	Dataset	classification method	Criteria
[8]	2024	Data cleaning, data scaling, and feature selection	UNSW-NB15	CNN + LSTM	Accuracy=99%
[9]	2022	Apache Spark and feature selection	UNSW-NB15	RF	Accuracy=97.49%
[10]	2018	N/A	UNSW-NB15	RF	Accuracy=97.49% Sensitivity=93.53%
[11]	2020	Data cleaning, data scaling, and feature selection	UNSW-NB15	DT	Accuracy=90.85%
[12]	2019	Normalization data	UNSW-NB15	SVM	Accuracy=85.99%
[13]	2019	Data cleaning	UNSW-NB15	Feed-Forward Neural Network	Accuracy=99.5%
[14]	2022	Data cleaning and data scaling	UNSW-NB15	Extra Tree	Accuracy=96.24%
[15]	2020	Feature selection	UNSW-NB15	CART	Accuracy=95.37% False acceptance rate= 11.86%
[16]	2021	Normalization data	UNSW-NB15 and CICIDS2017	NB+SVM	Accuracy=93.75 % for UNSW-NB15 and 98.92% for CICIDS2017
[17]	2021	Data cleaning, normalization, and feature selection	UNSW-NB15	Extreme Learning Machines	Accuracy=99%
[18]	2020	Normalization and One Hot Encoding	UNSW-NB15 and NSL-KDD	CNN-BiLSTM	Accuracy=82.08% for UNSW-NB15 and 99.22% for NSL-KDD

[19]	2023	Normalization	UNSWNB15	LSTM	Accuracy=99%

3. Methodology

The proposed framework methodology involves several key steps to ensure effective evaluation and optimization of classifiers. Initially, the UNSW-NB15 dataset undergoes cleaning to remove null values, followed by Z-score normalization to standardize the data. Feature selection is then performed using the Flower Pollination Optimization (FPO) to identify the best subset of features. The dataset is split into training (70%) and test (30%) sets, and various classifiers (MLP, LDA, QDA, KNN) are trained on the training set, as shown in Fig.1. Finally, the trained models are assessed using five metrics: confusion matrix, precision, F1-score, recall, and accuracy.

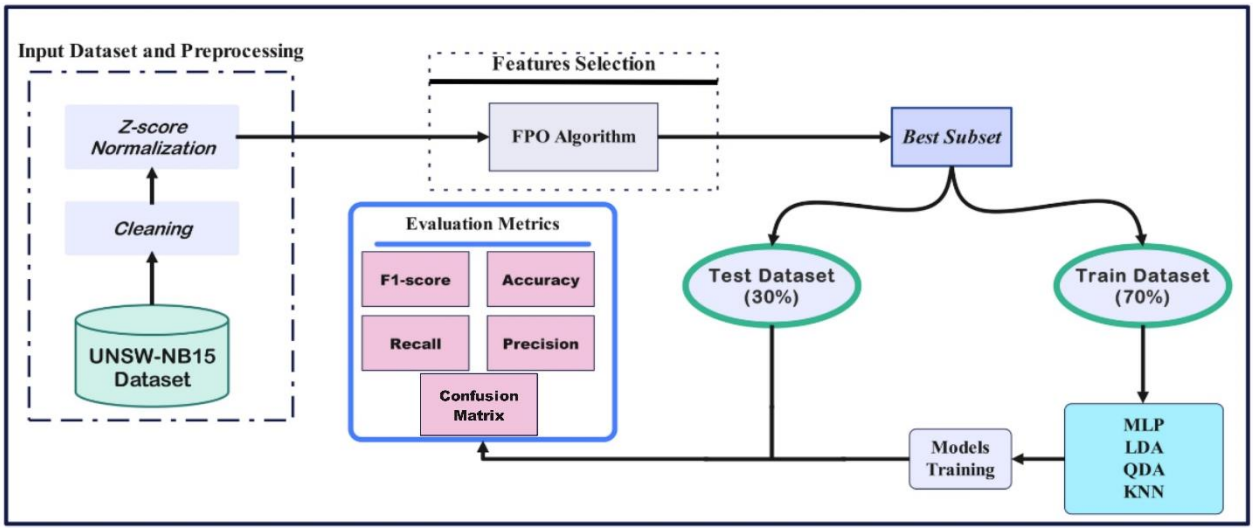

Fig. 1- Proposed framework methodology.

3.1. Dataset

In this paper, the suggested approach is evaluated and trained using the UNSWNB15 dataset. In this dataset, captured abnormal and normal IoT device traffic are included. The abnormal traffic helps identifying the potential attacks on IoT devices and the corresponding networks [20]. It has Bro, Argus, pcap, and CSV files of various data formats. The pcap file is the raw material or primary data source for network characteristic analysis. It is then developed into CSV files through Argus and Zeek IDS. Every record is labeled as either normal or an attack, all adding up to 45 attributes. The attacks are classified as follows: nine classes that weave in with each other, having something to do with security in IoT devices and network-based attacks, including Worms, Reconnaissance, Exploits, Backdoors, Fuzzers, Shellcode, Generic, DoS, and Analysis. This dataset is the most used one in packet analysis in IDS systems. The data set contains 257,673 rows, and how many rows are occupied by each category is captured in Table 2. The UNSWNB15 logs network traffic data in a controlled laboratory setup, viewing anomalies and patterns that may be pertinent in the context of IoT scenarios. Consequently, this would be very useful in the case of network-based attack analysis against IoT devices. Moreover, it is also suitable for evaluation and training of anomaly detection and IDS, which constitute a very important part of IoT security.

Table 2 - Description of UNSW-NB15 dataset.

Class	Testing Set	Training Set
Worms	44	130
Reconnaissance	3496	10,491
Exploits	11,132	33,393
Backdoors	583	1746
Fuzzers	6062	18,184
Shellcode	378	1133

Generic	18,872	40,000
DoS	4089	12,264
Analysis	677	2000
Normal	37,000	56,000

3.2. Preprocessing Dataset

It is the most important part of the analysis of data, more so in machine learning tasks, since it ensures a dataset that is clean, consistent, and appropriate to be used in training the model. Proper preprocessing will raise the efficiency of machine learning models to very great degrees by decreasing noise, handling missing values, and normalizing data this latter point makes sure that all features are on the same scale and have equal contributions to the model. First, this preprocessing step applied to the UNSW-NB15 dataset cleans up the dataset. There are no null or empty values; this will ensure that these null values often lower the accuracy obtained in a Machine Learning model and may alter an analysis. By treating those missing values, we can hold that our dataset is complete and the models trained on it can fully be trusted [21].

Normalization is an important step in preprocessing. This is more of a requirement of many machine-learning algorithms, particularly neural networks and k-nearest neighbors, which are sensitive to the scale of data. This can guarantee that every feature is contributing equally to the model and not by letting features with larger ranges dominate the features with a small range [22]. We implement here the Z-score normalization method. It rescales data onto a common scale with a mean of 0 and a standard deviation of 1. This method becomes very strong, especially when the data is normally distributed. The Z-score normalization of the feature X may be calculated using the following equation:

$$\boldsymbol{Z} = \frac{x - u}{\sigma} \tag{1}$$

Where z is the Z-score normalized value, x is the original value, u is the mean of the feature, σ is the standard deviation of the feature. This will standardize the dataset after applying the Z-score normalization, which can be more appropriate for algorithms of machine learning. It will make all features in the same scale, hence improving the model's performance and convergence.

3.3. Feature Selection based on FPO

Feature selection remains one of the most important stages in the pipeline of data preprocessing, since it fundamentally affects both the performance and the efficiency of machine learning models [23]. Key objectives for feature selection are to select or identify only those features in a given dataset relevant toward the variable of prediction or output. This will help reduce dimensions, eliminate redundant and irrelevant features, and increase the accuracy in the model with reduced overfitting. By ensuring that the model picks only relevant features, performance and predictive accuracy can be driven into focusing on the most critical elements of the data. It also reduces the model complexities—avoiding irrelevant and redundant features puts the breaks on overfitting. This also reduces training time. In the end, feature selection also helps get simpler models that are easy to interpret and maintain.

FPO stands for Flower Pollination Optimization, a new optimization process mimicking the pollination process of flowering plants [24]. The FPO technique has been successfully applied for feature selection in several datasets, such as UNSW-NB15, to determine the best subset of features that give maximum performance to models. FPO mimics the pollination process, wherein the pollen carriers pollinators, such as bees and birds have a certain possibility of moving to another flower after carrying pollen and depositing it, thus cross-pollinating for reproduction with variations. This optimization technique has two phases: global pollination (exploration phase) and local pollination (exploitation phase). The reason for choosing FPO over other optimization techniques is its unique adaptability and efficiency. FPO's ability to balance global exploration and local exploitation allows it to effectively navigate complex search spaces, identifying the most relevant features that enhance model performance and accuracy. These attributes make FPO a compelling choice for feature selection, ensuring that the most critical elements of the data are utilized for optimal machine learning model performance. The algorithm iteratively updates the position of pollen to converge to a better solution. Equations 1 and 2 mathematically model FPO. Equation 2 corresponds to global pollination, and Equation 3 corresponds to local pollination.

$$\boldsymbol{X(t+1)} = X(t) + L \times (X^*(t) - X(t)) \quad (2)$$

$$\boldsymbol{X(t+1)} = X(t) + \in \times (X_j(t) - X_k(t)) \quad (3)$$

Where $X(t)$ is the current position vector, $X^*(t)$ is the position vector of the best solution obtained so far, L is a step size drawn from a Lévy distribution, $\in$ is a random number in [0, 1], $X_j(t)$ *and* $X_k(t)$are different position vectors chosen randomly from the population, and t is the current iteration. In this paper, a focus is placed on the use of FPO for improving feature selection in anomaly and intrusion detection systems for IoT devices.

3.4. Evolution Metrics

Metrics are used to assess a machine learning model's performance, which is crucial to its efficacy and reliability. We will focus on five major measures in model evaluation: accuracy, recall, precision, F1 score, and confusion matrix.

- Accuracy: Perhaps the most straightforward of the evaluation metrics is simply accuracy, which is defined as the number of correctly classified examples within total instances in the dataset [25]. This will perhaps provide a general feel for the performance of the model and can be inadequate only by itself in scenarios like highly imbalanced datasets.

$$\boldsymbol{Accuracy} = \frac{TP + TN}{TP + TF + FN + TN} \quad (4)$$

- Recall: Recall denotes how well the model is at getting positive instances right, otherwise also called sensitivity or true positive rate. This becomes of great importance in situations when missing positives has high costs.

$$\boldsymbol{Recall} = \frac{TP}{TP + FN} \quad (5)$$

- Precision: Precision, also referred to as positive predictive value, measures how accurate the positive predictions are provided by this model. It represents the proportion of true positive cases among all the cases that the model has classified as positive.

$$\boldsymbol{Precision} = \frac{TP}{TP + FP} \quad (6)$$

- F1-Score: The F1-score is the harmonic mean of recall and precision, which gives a single metric that can balance these two metrics.

$$\boldsymbol{F1 - Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

- Confusion Matrix: The confusion matrix provides a concrete scenario of how well the model works by presenting the actual versus predicted classifications. It is a table with the following four entries: false negatives, true negatives, true positives, and false positives. A confusion matrix provides clarity on the type of mistakes your model is making and is quite a useful tool to evaluate classification performance. The four elements of the confusion matrix are shown in Fig. 2.

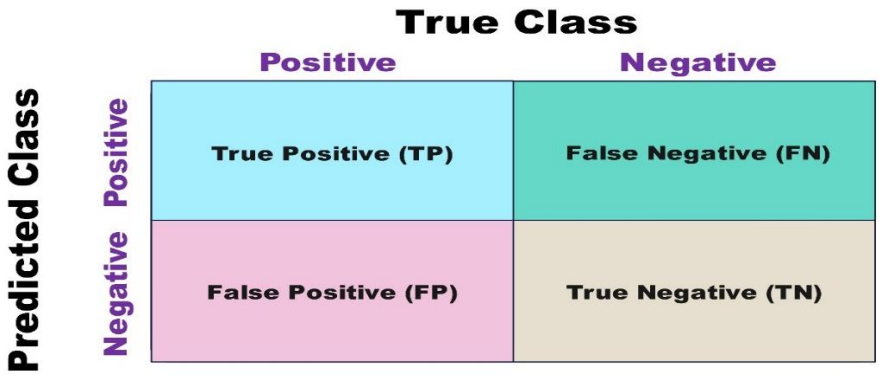

Fig. 2- Confusion matrix [26].

Each of these metrics contributes differently toward the model's performance. All these metrics looked at together may go a long way in bringing out a nuanced picture comprehensively of our anomaly and intrusion detection systems. This would enable fine-tuning and improving our models for better security in IoT environments.

4. Results and Discussion

In this section, the performances of the four classifiers (MLP, LDA, QDA, and KNN) are evaluated. The respective performances of the mentioned classifiers are assessed for five evaluation metrics: confusion matrix, recall, F1-score, precision, and accuracy. The above-mentioned evaluation is performed for two scenarios: first, when there is no feature selection, and second, when there is feature selection using the FPO. This comparison helps bring out the effect of feature selection on the classifier's performance.

4.1. Classification performance without feature selection

In this section, we evaluate the classifiers based on the five metrics without any feature selection as shown in Table 3.

Table 3 - Classifiers performance without FPO feature selection.

Classifier	Precision	F1_Score	Recall	Accuracy
MLP	86.55	86.29	86.04	90.06
LDA	87.68	91.07	94.74	93.79
QDA	90.14	91.35	92.6	93.83
KNN	87.7	86.23	84.8	89.92

It had a precision of 86.55, an F1-score of 86.29, a recall of 86.04, and an accuracy of 90.06. These metrics, obtained with this model, prove that although the MLP performs reasonably well, it fails to hit the highest level of performance among the classifiers tested. Its balanced precision and recall show that MLP is relatively good at correctly identifying both positive and negative instances, but there is room for improvement. LDA went very well with a precision of 87.68, an F1-score of 91.07, a recall of 94.74, and accuracy of 93.79. LDA was outstanding; it had the highest recall, thus proving that it is really great at correctly classifying positive instances. The high F1-scores and accuracy also show that without feature selection, it is good at sorting out this data, thus proving this as one of the best classifiers on this model.

In most of these main metrics, the QDA outperformed other classifiers. The QDA has shown high classification performance at 90.14 precision, with an F1-score of 91.35, the recall is 92.6, and 93.83 in accuracy. The high value of precision and the F1-score actually mean that QDA is good not only at classifying positive examples in the data set but also at having the smallest proportion of falsely marked negative instances. This renders QDA the most efficacious classifier without feature selection for the dataset. The KNN classifier resulted in the precision of 87.7, F1 of 86.23, recall of 84.8, and accuracy of 89.92. KNN functions acceptably but has the lowest recall and F1 score of the

classifiers; this is indicative that KNN will perform more poorly due to false negatives when testing out on these methods. It is also less on the accuracy end, indicating that KNN may not be effective in dealing with the entire feature set. Fig. 3 shows the confusion matrices for the classifiers without FPO feature selection. These matrices provide a detailed view of the performance of each classifier by displaying the true positive, true negative, false positive, and false negative counts. The matrices further illustrate the strengths and weaknesses of each classifier, complementing the summary statistics provided above.

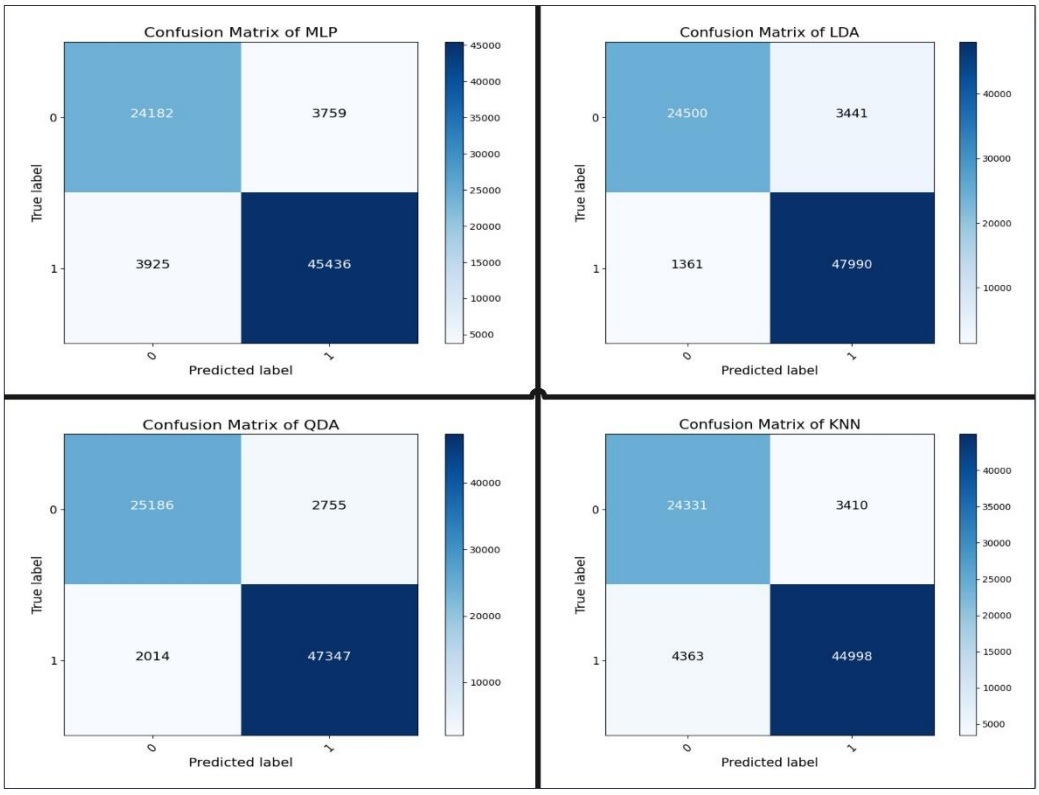

Fig 3- Confusion matrix of the classifiers without feature selection.

4.2. Classification performance with FPO feature selection

In this section, we evaluate the classifiers based on the five metrics with feature selection using the FPO algorithm as shown in Table 4.

Table 4 - Classifiers performance with FPO feature selection.

Classifier	Precision	F1_Score	Recall	Accuracy
MLP	96.03	96.35	96.67	97.37
LDA	98.54	98.64	98.74	99.02
QDA	98.1	98.83	99.59	99.16
KNN	95.1	95.17	95.25	96.45

With FPO feature selection, the MLP classifier significantly improves its performance. It achieved a precision of 96.03, a recall of 96.67, an F1-score of 96.35, and an accuracy of 97.37. These metrics spell out an enormous improvement in the capability of the right classifier to identify the positive and negative cases correctly. This meter is a clear demonstration of FPO's effectiveness in performance improvement for MLP. LDA also performed very great with the feature selection of FPO. An achieved precision of 98.54, an F1-score of 98.64, a recall of 98.74, and accuracy of 99.02%. These near-perfect metrics indicate the improved ability of LDA to correctly classify the instances—almost with no mistakes at all—proving the highly positive effect of feature selection on its performance.

QDA did very well with FPO feature selection. In this regard, an F1-score of 98.83, a precision of 98.1, with a recall of 99.59 and an accuracy of 99.16, were recorded for the QDA. The high recall and accuracy of the QDA indicate very good performance in identifying true positive instances with low false positives, hence the best classifying algorithm in this case. The KNN classifier also gained with FPO feature selection to an F1-score of 95.17, a precision of 95.1, a recall of 95.25, and an accuracy of 96.45. Although this is a huge gain in performance for KNN, its performance remained only marginally inferior to the other classifiers after FPO selection, indicating that FPO is particularly effective for this classifier but that KNN is not as truly effective as the discriminative analysis methods on this dataset.

Fig. 4. shows the confusion matrix of the classifiers after used FPO feature selection. The confusion matrix for the MLP classifier shows a significant improvement in performance after feature selection. The true positives (48,436) and true negatives (26,831) are high, while the false positives (1,110) and false negatives (925) are considerably reduced. This indicates that FPO has enhanced MLP's ability to correctly classify instances, substantially reducing errors compared to the scenario without feature selection. The confusion matrix for LDA reveals near-perfect classification results. LDA achieved 48,998 true positives and 27,539 true negatives, with only 402 false positives and 352 false negatives. These results demonstrate LDA's exceptional performance, with the lowest number of classification errors among all classifiers. More importantly, high true positives and true negatives directly give the high increase in accuracy and reliability due to FPO feature selection.

The QDA classifier's confusion matrix shows outstanding results, with 27,408 true negatives, and 49,247 true positives. The false positives (533) and false negatives (114) are minimal, highlighting QDA's superior ability to correctly classify instances. QDA the best-performing classifier with feature selection, demonstrating its enhanced effectiveness in reducing both types of classification errors. The KNN classifier also shows notable improvements in its confusion matrix. KNN achieved 47,998 true positives and 27,331 true negatives, with a reduction in false positives (1,410) and false negatives (1,363). Although its performance is slightly lower than that of QDA and LDA, KNN still shows substantial improvements, indicating the positive impact of FPO feature selection in enhancing its classification performance. The analysis of the confusion matrices after performing FPO feature selection highlights substantial performance improvements across all classifiers. QDA and LDA emerge as the top-performing classifiers, with QDA achieving the highest number of true positives and the lowest number of false negatives. LDA also demonstrates near-perfect classification results with the lowest number of false positives and false negatives. MLP and KNN show significant improvements in reducing classification errors, further illustrating the effectiveness of FPO in enhancing classifier performance.

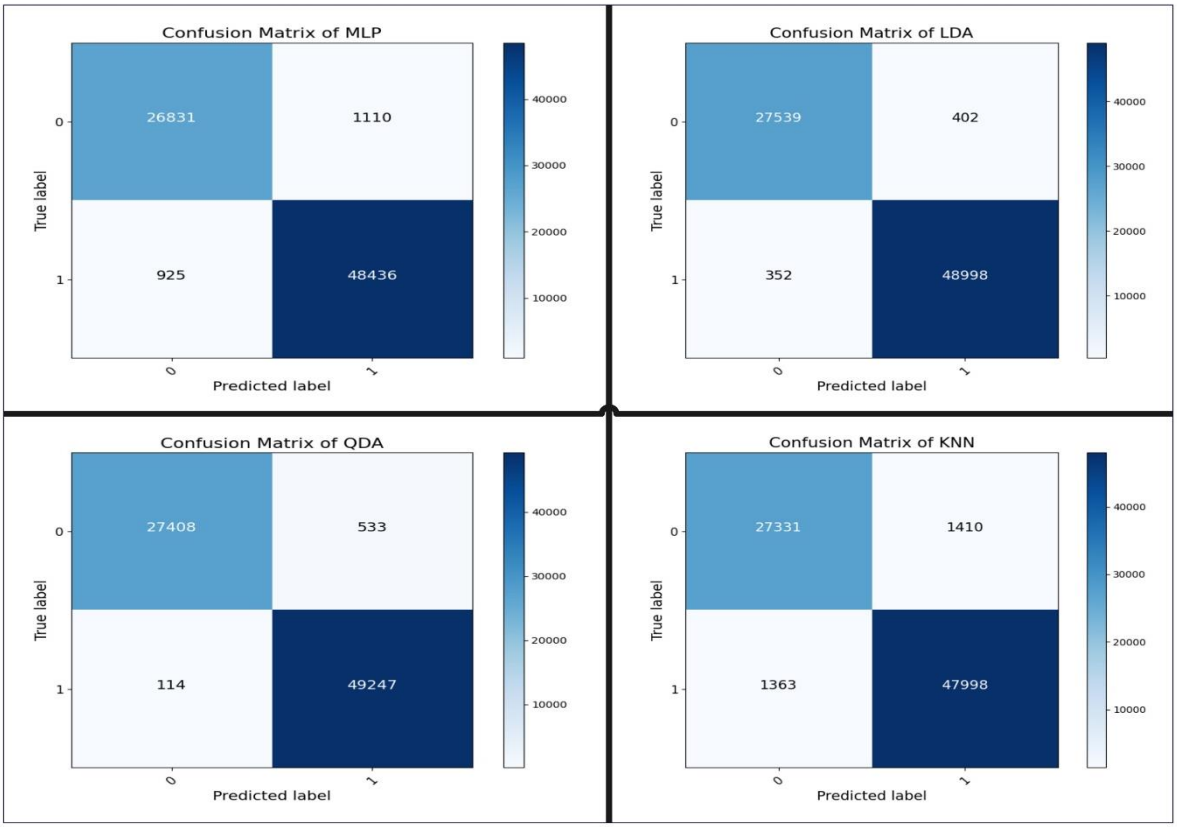

Fig 4- Confusion matrix of the classifiers with FPO feature selection.

4.3. Comparison with recent studies

In this section, we compare the performance of our proposed method with recent studies that have used the UNSW-NB15 dataset. The comparison focuses on the accuracy achieved by different classifiers and feature selection techniques as shown in Table 5.

Table 5 - Comparison with recent studies on the UNSW-NB15 dataset.

Ref	Year	Feature Selection	Classifier	Accuracy (%)
[11]	2020	XGBoost	DT	90.85
[9]	2022	PSO	RF	97.49
[14]	2022	Mutual Information Gain feature selection	Stacking machine learning models	96.24
[27]	2024	Random sampling and correlation analysis	RF	98.63
Our Proposed		**FPO**	**QDA**	**99.16**

As shown in Table 5, various feature selection methods have been employed in recent research, including XGBoost, PSO, Mutual Information Gain, and correlation analysis. XGBoost has been frequently applied across different models. Classification techniques have involved both machine learning models like Random Forests and Decision Trees, as well as advanced ensemble methods. Although Random Forests achieved high accuracy, the highest reported accuracy in previous studies was 98.63%, using random sampling and correlation analysis with Random Forests. In comparison, our proposed method using the FPO algorithm for feature selection has demonstrated superior performance. In particular, our approach appreciated accuracies up to 99.16% with the QDA classifier and 99.02% with the LDA classifier. Furthermore, the MLP and KNN classifiers also produced better results, such as 97.37% accuracy and 96.45%, respectively. The results indicate that our suggested approach outperformed all the previously reported approaches concerning accuracy and thus set a new benchmark on anomaly detection and intrusion detection systems within IoT environments.

5. Conclusion

This paper presents an investigation into the Flower Pollination Optimization (FPO) algorithm for feature selection to enhance the performance of various classifier models trained on the UNSW-NB15 dataset. The preprocessing included cleaning the dataset and normalization using Z-score. Four classifiers were evaluated: Multi-Layer Perceptron (MLP), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and K-Nearest Neighbors (KNN), under two scenarios: without feature selection and with FPO-based feature selection. The results indicate that feature selection using FPO significantly improved the performance of all classifiers. Specifically, QDA and LDA achieved accuracies of 99.16% and 99.02%, respectively, while MLP and KNN showed major improvements, with accuracies reaching 97.37% and 96.45%, respectively. Comparisons with recent studies demonstrated that the proposed method outperformed existing approaches, setting a new benchmark in accuracy for anomaly detection and intrusion detection systems in IoT environments. The superior performance of the method highlights the importance of effective feature selection for improving the accuracy and reliability of machine learning models. The FPO-based feature selection methodology proved highly effective in enhancing the performance of classifiers on the UNSW-NB15 dataset. This approach can be highly useful for implementing various applications related to anomaly detection, providing better robustness and efficiency in IoT cybersecurity applications. Future work should test this feature selection technique on more real-life datasets and classifiers, and integrate it into real-time detection systems.

Acknowledgements

We would like to express our gratitude to all the individuals and institutions who supported and contributed to this research. Special thanks to the College of Computer Science and Information Technology at the University of Al-Qadisiyah for their invaluable assistance and resources that made this study possible.

References

[1] H. Liu and B. Lang, "Machine learning and deep learning methods for intrusion detection systems: A survey," applied sciences, vol. 9, no. 20, p. 4396, 2019.
[2] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," Ieee Access, vol. 7, pp. 41525–41550, 2019.
[3] M. Almseidin, M. Alzubi, S. Kovacs, and M. Alkasassbeh, "Evaluation of machine learning algorithms for intrusion detection system," presented at the 2017 IEEE 15th international symposium on intelligent systems and informatics (SISY), IEEE, 2017, pp. 000277–000282.
[4] A. Aldweesh, A. Derhab, and A. Z. Emam, "Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues," Knowledge-Based Systems, vol. 189, p. 105124, 2020.
[5] V. Kumar, D. Sinha, A. K. Das, S. C. Pandey, and R. T. Goswami, "An integrated rule based intrusion detection system: analysis on UNSW-NB15 data set and the real time online dataset," Cluster Computing, vol. 23, pp. 1397–1418, 2020.
[6] T. Saba, A. Rehman, T. Sadad, H. Kolivand, and S. A. Bahaj, "Anomaly-based intrusion detection system for IoT networks through deep learning model," Computers and Electrical Engineering, vol. 99, p. 107810, 2022.
[7] R. Abdulhammed, M. Faezipour, A. Abuzneid, and A. AbuMallouh, "Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic," IEEE sensors letters, vol. 3, no. 1, pp. 1–4, 2018.
[8] K. Psychogyios, A. Papadakis, S. Bourou, N. Nikolaou, A. Maniatis, and T. Zahariadis, "Deep Learning for Intrusion Detection Systems (IDSs) in Time Series Data," Future Internet, vol. 16, no. 3, p. 73, 2024.
[9] R. Tahri, A. Jarrar, A. Lasbahani, and Y. Balouki, "A comparative study of Machine learning Algorithms on the UNSW-NB 15 Dataset," presented at the ITM Web of Conferences, EDP Sciences, 2022, p. 03002.
[10] M. Belouch, S. El Hadaj, and M. Idhammad, "Performance evaluation of intrusion detection based on machine learning using Apache Spark," Procedia Computer Science, vol. 127, pp. 1–6, 2018.
[11] S. M. Kasongo and Y. Sun, "Performance analysis of intrusion detection systems using a feature selection method on the UNSW-NB15 dataset," Journal of Big Data, vol. 7, no. 1, p. 105, 2020.
[12] D. Jing and H.-B. Chen, "SVM based network intrusion detection for the UNSW-NB15 dataset," presented at the 2019 IEEE 13th international conference on ASIC (ASICON), IEEE, 2019, pp. 1–4.
[13] L. Zhiqiang, G. Mohi-Ud-Din, L. Bing, L. Jianchao, Z. Ye, and L. Zhijun, "Modeling network intrusion detection system using feed-forward neural network using unsw-nb15 dataset," presented at the 2019 IEEE 7th International Conference on Smart Energy Grid Engineering (SEGE), IEEE, 2019, pp. 299–303.
[14] M. H. Kabir, M. S. Rajib, A. S. M. T. Rahman, M. M. Rahman, and S. K. Dey, "Network intrusion detection using unsw-nb15 dataset: Stacking machine learning based approach," presented at the 2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), IEEE, 2022, pp. 1–6.
[15] Z. Chkirbene, S. Eltanbouly, M. Bashendy, N. AlNaimi, and A. Erbad, "Hybrid machine learning for network anomaly intrusion detection," presented at the 2020 IEEE international conference on informatics, IoT, and enabling technologies (ICIoT), IEEE, 2020, pp. 163–170.
[16] J. Gu and S. Lu, "An effective intrusion detection approach using SVM with naïve Bayes feature embedding," Computers & Security, vol. 103, p. 102158, 2021.
[17] S. Moualla, K. Khorzom, and A. Jafar, "Improving the Performance of Machine Learning-Based Network Intrusion Detection Systems on the UNSW-NB15 Dataset," Computational Intelligence and Neuroscience, vol. 2021, no. 1, p. 5557577, 2021.
[18] J. Sinha and M. Manollas, "Efficient deep CNN-BiLSTM model for network intrusion detection," presented at the Proceedings of the 2020 3rd International Conference on Artificial Intelligence and Pattern Recognition, 2020, pp. 223–231.
[19] J. Jose and D. V. Jose, "Deep learning algorithms for intrusion detection systems in internet of things using CIC-IDS 2017 dataset," International Journal of Electrical and Computer Engineering (IJECE), vol. 13, no. 1, pp. 1134–1141, 2023.
[20] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," presented at the 2015 military communications and information systems conference (MilCIS), IEEE, 2015, pp. 1–6.
[21] A. Palanivinayagam and R. Damaševičius, "Effective handling of missing values in datasets for classification using machine learning methods," Information, vol. 14, no. 2, p. 92, 2023.
[22] M. Wu et al., "A fault detection method of electric vehicle battery through Hausdorff distance and modified Z-score for real-world data," Journal of Energy Storage, vol. 60, p. 106561, 2023.
[23] M. Noaman Kadhim, D. Al-Shammary, and F. Sufi, "A novel voice classification based on Gower distance for Parkinson disease detection," International Journal of Medical Informatics, vol. 191, p. 105583, Nov. 2024, doi: 10.1016/j.ijmedinf.2024.105583.
[24] D. Rodrigues, X.-S. Yang, A. N. De Souza, and J. P. Papa, "Binary flower pollination algorithm and its application to feature selection," Recent advances in swarm intelligence and evolutionary computation, pp. 85–100, 2015.
[25] M. N. Kadhim, A. H. Mutlag, and D. A. Hammood, "Multi-models Based on Yolov8 for Identification of Vehicle Type and License Plate Recognition," presented at the National Conference on New Trends in Information and Communications Technology Applications, Springer, 2023, pp. 118–135.
[26] D. Al-Shammary, M. N. Kadhim, A. M. Mahdi, A. Ibaida, and K. Ahmed, "Efficient ECG classification based on Chi-square distance for arrhythmia detection," Journal of Electronic Science and Technology, vol. 22, no. 2, p. 100249, 2024.
[27] S. More, M. Idrissi, H. Mahmoud, and A. T. Asyhari, "Enhanced Intrusion Detection Systems Performance with UNSW-NB15 Data Analysis," Algorithms, vol. 17, no. 2, p. 64, 2024.