Journal of Innovative Construction and Petrochemical Technologies, Vol.3, No.2 (December, 2026), 88-103

P-ISSN: xxxx- xxxx | E-ISSN: 3007-9861 | DOI: https://doi.org/10.71285/icpt.v3i2.34

Journal of Innovative Construction and Petrochemical Technologies
https://icpt.ouk.kz/index.php/icpt/index

Research Article

Check for updates

CDDO-Leo-Based Feature Selection for IoT Intrusion Detection

Mustafa Ahmed Abdulwahhab
Department of Administrative and Financial Affairs, Al-Iraqia University, Baghdad, Iraq

Abstract: An Internet of Things (IoT) intrusion detection system (IDS) can be defined as one of the security systems designed for monitoring, analyzing, and detecting malicious cyberthreats or activity on IoT network. It helps to protect IoT devices as well as infrastructure from malware, unauthorized access, denial-of-service (DoS) attacks, and other security breaches. IoT networks produce high-dimensional, redundant and noisy data and thus intrusion detection is a difficult task. The current techniques like HHO-SSA, GWO, and WOA are limited in terms of balancing exploitation and exploration, which results in sub optimal feature selection and low classification performance. To enhance classification accuracy and computational efficiency, the proposed CDDO-LEO hybrid algorithm enhances the feature selection through integrating Child Drawing Development Optimization algorithm (CDDO) with Lagrange Elementary Optimization algorithm (LEO) for dataset refinement. It was used to test an IoT intrusion detection dataset where 17 optimal features were selected and Gradient Boosting got 98.60% accuracy. It performed better than the conventional feature selection methods like HHO-SSA, GWO and WOA. The comparative analysis establishes the fact that CDDO-LEO has better and more consistent classification performance especially when using ensemble classifiers. The proposed method also consumes less processing time and is also has high accuracy thus useful in high dimensional datasets.

Keywords: Feature Selection, Machine Learning, Intrusion Detection System (IDS), Cybersecurity, IoT Security

INTRODUCTION

The IoT is changing the world of industries very fast, as it allows facilitation of seamless connectivity among smart devices. These interconnected systems produce enormous volumes of data, which push the development of automation, medical, smart cities and industry. Nevertheless, the IoT networks are very exposed to cyber threat because of their massive implementation, lack of sufficient computational resources, and the absence of standardized security protocols [1].

Machine learning is a field that has great demands in efficiency with regard to processing and analysis of big data due to its rapid evolution. A feature selection is a vital preprocessing data transformation step that is used for dimension reduction that increase the system's predictive capabilities. With the use of feature selection, the most significant features are selected, resulting in better model interpretation and reduced computations because it will reduce the likelihood of model overfitting [2].

Security is an important consideration since malicious attacks include unauthorized access, malware, denial-of-service (DoS) attacks and data breaches, all of which are major threats of IoT infrastructures. IDS are one of the extremely important defense systems that have been established to protect the IoT systems against cyber threats. IDS monitor on network traffic, analyze system activity and detect malicious activity.

However, traditional IDS solutions face significant challenges, such as high-dimensional, redundant, and noisy data, which makes accurate intrusion detection complex. Additionally, conventional rule-based IDS struggle to adapt to new and evolving attack patterns, while signature-based methods can only detect known threats, leaving zero-day vulnerabilities undetected [3]. Wrapper-based methods excel at feature selection through predictive modeling approaches because they achieve superior accuracy rates according to research findings by evaluating subsets of features. However, fast and practical methods must be created due to the computational complexity of current approaches [4]. The selection of features through traditional evaluation methods depends on statistical measures to determine relevancy without regard for model predictive capabilities. Such methods show computational efficiency, but they do not effectively discover complex feature combinations. The learning algorithm implements feature selection within its structure during embedded methods, though these methods tend to work only with certain modeling frameworks. The search process for optimal feature subsets using wrapper-based methods requires multiple predictive model training to identify optimal feature subsets at high accuracy, and it also requires more computational power [5].

The contribution of this paper is to introduces a hybrid feature selection approach which combines two modern metaheuristic optimization algorithms, namely Child Drawing Development Optimization algorithm (CDDO) and Lagrange Elementary Optimization algorithm, (LEO) [6] [5]. CDDO draws its inspiration from the drawing processes of child cognitive development to enable broad-based and exploratory search operations. Through swarm-based optimization, LEO implements leader-follower dynamics, which locates important features by focusing on local exploitation. The hybrid CDDO-LEO method integrates two search methods to reach the best balance between discovery-oriented exploration and focused exploitation of relevant features [7].

The following are the main goals of the presented study:

1. Creating a new hybrid feature selection algorithm that combines the exploitative strengths regarding LEO with the CDDO's exploratory strengths.
2. Assessing the hybrid CDDO-LEO algorithm's performance utilizing benchmark datasets with regard to regression and classification tasks.
3. Evaluating the suggested algorithm in comparison to standalone LEO, CDDO, and other traditional feature selection methods.
4. Evaluating the hybrid algorithm's robustness, scalability, and computing efficiency on datasets with different feature dimensionalities.

This study is of great importance because it develops the feature selection method through the combination of improved accuracy and efficiency. Machine learning applications in bioinformatics, financial modeling, image analysis and medical diagnostics needs a stable feature selection method for managing the data with large features [8]. LEO with CDDO functionality will offer flexible, optimized system beyond the existing methods at the same time maintaining the practical computational efficiency.

Future studies will be enriched by intelligent use of complementary nature of two optimization strategies in intelligent feature selection and optimization strategies. The hybrid approach of CDDO-LEO uses fewer features than before while maintaining accuracy and it allows machine learning models to have more interpretability and speedup regarding computational resources and have a higher level of generalization capability [10].

The principal originality of the research is the development and use of the CDDO-LEO hybrid algorithm in feature selection in high-dimensional IoT intrusion datasets. The proposed approach, as opposed to traditional sequential hybrids, combines exploration (CDDO) and exploitation (LEO) in a structured phase-based approach, so that the optimal search of feature subsets can be balanced.

Furthermore, it was strictly tested on extensive IoT datasets with more than 1.1 million cases as well as a variety of classifiers, and it has outperformed existing and standalone metaheuristic techniques with regard to dimensionality reduction, classification accuracy, and—majorly—computational efficiency. The work's hybrid-based design, mass application, and thorough empirical validation set it apart from other researchers' work and provide useful benefits for IoT security systems in real world. This study's remaining sections are organized as follows: The literature review of relevant work in intrusion detection as well as feature selection is presented in Section 2. Section 3 describes the suggested CDDO-LEO technique. The experimental design, findings, and comparative analysis are covered in Section 4. Finally, Section 5 provides future study directions as a summary of the study.

LITERATURE REVIEW

Machine learning studies also paid much attention to feature selection by conducting numerous studies on the approach to selecting the best feature selection techniques to preserve the accuracy of the prediction. Recent developments that optimize the techniques of feature selection have been enhanced by swarm intelligence in conjunction with evolutionary computation and hybrid techniques [11]. The feature selection field implements two common approaches of mathematical algorithms, i.e. Genetic Algorithms (GA) and Particle Swarm Optimization (PSO).

Zaman et al. (2022) research resulted in an optimized PSO approach that is effective on large datasets as it enhances the accuracy of the measurement and accelerates the computation [11].

Luan and Thinh (2023) introduced a hybrid GA-based feature selection model that enhanced the performance of the medical imaging applications in classification in an effective way [12].

These methods are faced with premature convergence and exploration and exploitation discrepancies. Researchers in modern times have adopted hybrid optimization techniques in order to eliminate the existing limitations of optimization techniques. Widians et al. (2022) developed a hybrid algorithm, which incorporated the use of the Ant Colony Optimization (ACO) and Grey Wolf Optimizer (GWO) for feature selection and achieve a better classification performance compared to the previously obtained [13].

Similarly, Gupta and Kumar (2023) have developed a hybrid framework integrating Harris Hawks Optimization (HGO) with a Deep Learning-based classifier, which enhanced significantly the efficiency of features selections in the analysis of a financial time-series analysis [14]. Other than these enhancements, most of the hybrid approaches are computationally intensive and also demand the optimization of many hyperparameters.

A recent application of CDDO algorithm has been used as a potential metaheuristic method, which is based on cognitive learning patterns among children. Scholars like Abdulhameed and Rashid (2022) have proven the capability of CDDO in feature selecting after it reached higher convergence rates, as well as exploration potential than traditional swarm intelligence algorithms [15].

Nonetheless, standalone CDDO has a problem of local optima tapping, and this requires additional improvement. Recent LEO innovation shows effective ability of feature selection through a combination of swarm intelligence and a leader-based exploitation strategy. Recent studies by Kristiyanti et al. (2023) applied LEO to large biomedical datasets to achieve the best classification results with greater efficiency when compared to traditional methods [16]. LEO, however, is not diverse in the initial search space, which is vulnerable to suboptimal feature subsets.

As an effort to address these shortcomings in separate algorithms, the paper suggests a hybrid algorithm CDDO-LEO that can integrate the ability of exploration of CDDO with the power of exploitation of LEO. These methods will result in the hybrid model being able to offer a more computationally effective and balanced feature selection method that will improve classification and regression performance across the many domains [17].

Table 1. Summary of Literature Review

Reference	Method Used	Limitations	Future Work Suggestions
Zaman et al. (2022)	Optimized PSO	Premature convergence, lacks stability in feature selection	Hybridizing PSO with another algorithm to improve feature subset stability
Luan & Thinh (2023)	Hybrid GA-based selection	Computationally expensive, requires tuning	Implementing adaptive GA to reduce parameter dependence
Widians et al. (2022)	GWO + ACO hybrid	High computational cost	Reducing execution time via distributed computing
Gupta & Kumar (2023)	HHO + Deep Learning	Overfitting risk, limited interpretability	Exploring explainable AI techniques in feature selection
Abdulhameed & Rashid (2022)	CDDO	Gets trapped in local optima	Hybridizing CDDO with an exploitation-focused optimizer
Kristiyanti et al. (2023)	LEO	Lacks diversity in search space	Enhancing diversity by combining LEO with an exploration-based algorithm
Proposed Work (2024)	CDDO + LEO hybrid	Balances exploration and exploitation effectively	Applying the technique to real-time IDS

METHOD

This section could be divided into the next subsections:

Classifiers

Classification aims to label newly encountered samples without assigning the classification. Before classification begins, it is necessary to train the classifier to understand data characteristics and identify the way attribute values as well as class labels. To evaluate selection performances, the research uses four classifiers in its methodology. The Decision Tree (DT) classifier finds broad acceptance because of its ability to generate easily interpretable results and quick computational capabilities. The model builds a hierarchical structure that uses nodes as decision rules for feature values leading to class label leaves. This tool demonstrates its best use in attribute selection and classification tasks as it reveals the importance level of each characteristic. Random Forest (RF) classifiers construct groups of decision trees and combine them for better accuracy along with the reduction of overfitting patterns. Random feature selection per split in RF improves both the features selection stability and generalization properties, which makes it an excellent solution for intrusion detection systems. K-nearest neighbors (KNN) classifier applies a non-parametric approach for data point classification through measuring the proximity between records in the feature space. The method remains straightforward while delivering results because the user needs only to specify the parameter. This parameter identifies the number of neighboring observations to use. The KNN method enables network traffic anomaly detection and allows identification of the most suitable feature set from the available attributes. Gradient boosting (GB) represents a persistent machine learning algorithm which constructs different models one by one by allowing successive models to refine previous incorrect outputs. High predictive accuracy features are the major characteristic of this technique that makes it suitable for supervised learning problems, including high-dimensional IoT intrusion detection systems.

Algorithms and Theories

CDDO

The algorithm CDDO operates through simulating three different developmental phases of children when they create drawings. The algorithm harnesses natural principles by using a set of equations analyzed from drawing behavior observations, allowing the algorithm to maintain exploration and exploitative operations. The CDDO search process depends on manual pressure changes and drawing sketches according to the golden ratio while keeping memory of past optimal solutions [18]. The mathematical formulation of CDDO follows these key equations:

$$RHP = LB + (UB - LB) \times rand(0,1) \rightarrow (1)$$

Relative hand pressure (RHP) to be used in drawing process is calculated using equation 1. The following values LB and UB that establish lower and upper limits of search space render the exploration reasonable. The function rand (0,1) adds the element of randomness to enable a variety of exploration with regard to feature subsets [19].

$$X_i^{t+1} = X_i^t + \varphi \times (L + W) \rightarrow (2)$$

Equation 2, in the same manner, transforms the positions of the candidate solutions in accordance with the golden ratio, φ, and is close to 1.618. The terms L and W indicate the length and the width of explored area, respectively. The golden ratio makes the CDDO offer a balance in the exploration and refinement which improves the rate of convergence.

$$X_i^{t+1} = X_i^t + CR \times \left(X_{best} - X_{memory}\right) \rightarrow (3)$$

The algorithm to learn from previously discovered optimal solutions will show in equation 3. The correction rate CRCR controls how much influence past best solutions X_{best} and stored memory solutions X_{memory} have on the current iteration. This mechanism helps avoid stagnation and enhances adaptive learning.

$$X_i^{t+1} = X_i^t + \alpha \times rand(-1,1) \rightarrow (4)$$

Here, α\alpha serves as a scaling factor, and rand (-1,1) introduces stochastic variation, preventing premature convergence to local optima in equation 4. This mutation process gives variation in feature selection therefore improving the strength of algorithm [19]. CDDO generates a population of candidate solutions, and a classifier (such as Decision Tree) is used to measure the fitness and iteratively updating the feature selection until convergence.

LEO

The LEO algorithm is based on swarm intelligence for achieving the Lagrange dynamics, that drives follower agents in adaptive leadership dynamics. The refinement of the feature selection is carried out through a series of position adjustments based on the fitness evaluation outcome [20]. Important LEO mathematical formulations are:

$$X_i^{t+1} = X_{leader} + \lambda (X_i^t - X_{leader}) \rightarrow (5)$$

Equation 5 updates the position of each agent Xi by moving it closer to the leader Xleader, who represents the best solution at current iteration. The parameter λ\lambda is an adaptive learning coefficient that dictates the degree of influence the leader has on the followers.

$$L(X) = f(X) + \mu \sum_{i=1}^{n} g_i(X) \rightarrow (6)$$

The Lagrange function is defined by Equation 6, and f(X) is the objective function (e.g., classification accuracy) and gi (X) is the constraints. The trade-off between optimizing feature selection and maintaining constraints is facilitated by the Lagrange multipliers μ.

$$X_i^{t+1} = X_i^t + \beta \times rand(-1,1) \times |X_{best} - X_i^t| \rightarrow (7)$$

Equation 7 has the adaptive perturbation mechanism to avoid the trapping of the agents in local minima. The perturbation factor β\beta modulates the magnitude of random changes depending on the distance between the current position of the agent and the best-known solution Xbest. This guarantees a more diversified process of search [20].

$$X_i^{t+1} = X_i^t + \gamma \times (X_{best} - X_{mean}) \rightarrow (8)$$

Equation 8 refines the solution space with adjustment of agent positions with respect to the global best Xbest and the population mean position Xmean. The convergence speed is governed by the parameter γ/gamma, so that in the subsequent iterations LEO will be concentrated on good solutions.

LEO starts with initial set of features, which gets enhanced until the best collection is discovered by employing the combination of the equations explained. LEO is a useful mechanism to CDDO because it enhances the capabilities of exploitation as well as convergence.

Combining LEO and CDDO produces a more enhanced hybrid selection system that can be optimized to increase both exploration and refinement to maximize the predictive performance in both classification and regression tasks. The combined system makes use of its capabilities to deliver computational efficiency as well as the selection of the relevant features.

CDDO-LEO

The algorithm of suggested CDDO-LEO hybrid feature selection combines the exploratory strengths of CDDO and the exploitative capabilities of LEO. The phases of the integration are organized as follows:

1. Initialization Phase: A population of feature subsets is randomly initialized.
2. Exploration Phase (CDDO): CDDO identifies promising feature subsets by feature selection.
3. Refinement Phase (LEO): LEO fine-tunes selected features by optimizing their inclusion probabilities.
4. Final Feature Selection: The best-performing subset is chosen based on classification/regression accuracy.

Feature selection is performed in a binary search space where each feature is represented as either 1 (selected) or 0 (excluded). The Binarization function converts continuous values into binary representation using a threshold [21]:

$$X_i^{bin} = \begin{cases} 1, if\ X_i > threshold \\ 0, otherwise \end{cases} \rightarrow (9)$$

where the threshold is a predefined parameter.

Fitness of the subset of features is assessed with the use of a classification/regression model. The objective function is determined as [22]:

$$Fitness = \alpha(1 - ClassificationError) + \beta\left(1 - \frac{|SelectedFeatures|}{TotalFeatures}\right) - (10)$$

where α and β control the balance between classification performance and feature reduction. Below are the details of Algorithm 1, presents the pseudocode of the proposed CDDO-LEO hybrid feature selection method. The algorithm integrates the exploration capability of CDDO with the exploitation strength of LEO to efficiently search for the optimal subset of features.

Algorithm 1: Hybrid CDDO-LEO Feature Selection.

Input: Population size PP, Iteration number TT, Lower bound LB=0LB = 0, Upper bound UB=1UB = 1, Correction rate CRCR, Scaling factor α

Output: Best selected feature subset F*

1. **Initialize** population of feature subsets Xi where i=1, 2, …, P
2. **Compute** initial fitness values of all feature subsets
3. **Set** Xbest as the best solution found so far
4. **While** (termination condition not met) do:

Evaluate fitness values of all feature subsets

1. **Apply** CDDO: - Compute RHP using Eq. (1)

Update feature positions using golden ratio (Eq. 2)

Adjust solutions based on pattern memory (Eq. 3)

2. **Apply** LEO: - Update agent positions using leader-follower strategy (Eq. 4)

Apply adaptive perturbation to avoid local optima (Eq. 5)

Refine features using exploitation strategy (Eq. 6)

3. **Select** the best subset based on fitness evaluation
   a. **End While**
   b. **Return** optimal feature subset F∗F^*

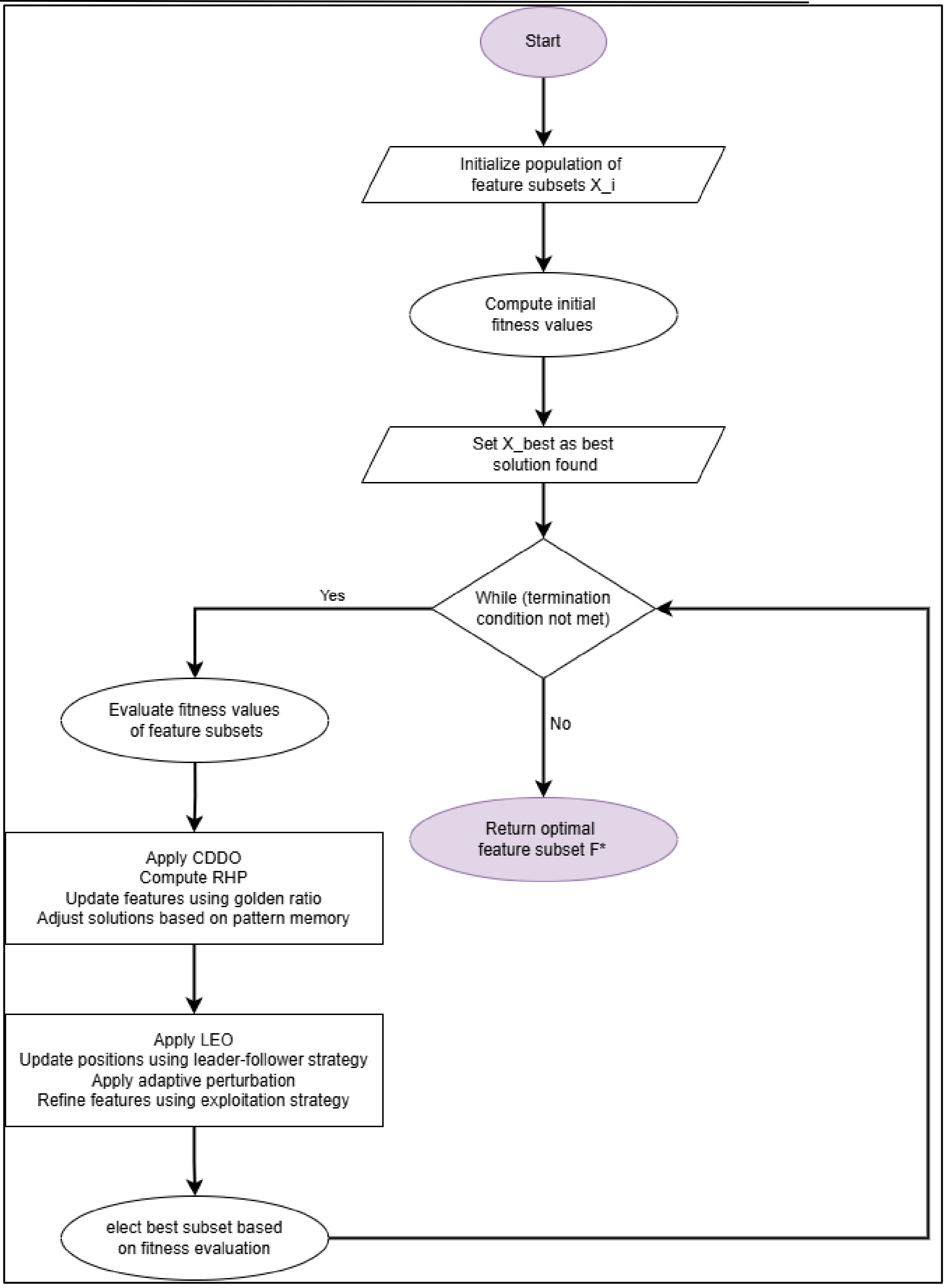

Figure 1. Structure of proposed CDDO-LEO

This structured pseudo-code will provide a comprehensive step-by-step explanation of the hybrid CDDO-LEO algorithm which ensures clarity and reproducibility. The justification of their hybridization as far as LEO and CDDO are concerned is based on their complementary optimization characteristics. CDDO possesses great abilities of global exploration and maintains population diversity which is beneficial in the high-dimensional feature space as well as avoiding early convergence. LEO is designed in the direction of the efficient local exploitation under the leader-follower dynamics and rapid convergence to the quality solutions. The existing hybrid techniques mostly combine optimizers with similar search characteristics, so they end up being redundant; however, LEO and CDDO were designed to be complementary. The hybrid suggested has an optimal trade-off of exploration-exploitation, making the convergence of the proposed hybrid more stable and feature selection.

Dataset

The current research is based on a dataset, which has various types of IoT intrusions. The dataset is structured in a manner that facilitates intrusion detection within IoT networks through network traffic data analysis. Each one of the intrusion types consists of multiple subcategories, and this makes the data even more complex and applicable in predictive modeling.

The dataset consists of 1,191,264 network traffic data instances, and each instance has 47 features which characterizes the various ways of the intrusion [23]. The types of IoT intrusions, which are found in the dataset, include:

1. DDoS (Distributed Denial of Service)
2. Brute-Force Attacks
3. Spoofing Attacks
4. Denial of Service (DoS) Attacks
5. Reconnaissance (Recon) Attacks
6. Web-based Attacks
7. Mirai-based Attacks

Performance Evaluation

In the presented work, the classification performance is measured using the following assessment metrics: precision, accuracy, F1, and recall.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \rightarrow (11)$$

$$precision = \frac{TP}{TP + FP} \rightarrow (12)$$

$$Recall = \frac{TP}{TP + FN} \rightarrow (13)$$

$$F1 = \frac{2 \times Specificity \times Recall}{Specificity + Recall} \rightarrow (14)$$

$$Specificity = \frac{TN}{TN + FP} \rightarrow (15)$$

The evaluation of classifier effectiveness in this study depends on True Positive (TP) and True Negative (TN) statistics, which indicate correct identifications of intrusions and normal traffic, respectively. The implementation of False Positive (FP) leads to the incorrect classification of normal traffic as intrusion while False Negative (FN) results from incorrect classification of intrusions as normal traffic. The evaluation of an IoT intrusion detection model depends heavily on these metrics to determine accuracy and precision as well as recall metrics which measure model performance when detecting intrusions.

RESULT AND DISCUSSION

The suggested hybrid CDDO-LEO algorithm has been trained using an IoT intrusion detection dataset to demonstrate its feature selection capabilities and classification performance. The dataset includes 1,191,264 network intrusion instances with 47 different features. After applying the CDDO-LEO hybrid algorithm, a total of 17 different features were selected, significantly reducing the dimensionality while preserving classification accuracy.

The dataset was divided into 80% of data used in training and 20% of data used in testing and various classifiers were applied to determine the efficiency of the selected features. The parameters of the four classifiers applied are introduced in Table 2 Decision Tree, Random Forest, KNN, Gradient Boosting.

Table 1. Parameter Values for Used Methods

Method	Parameter Values
CDDO-LEO Algorithm	Feature size: 47 Selected features: 17 Population size: 30 Iterations: 50 Correction rate (CR): 0.8
KNN Classifier	K = 5 Classes count: 2 Training set size: 80%
Decision Tree Classifier	Criterion: Gini Max depth: None Training set size: 80%
Random Forest Classifier	Number of trees: 100 Criterion: Gini Training set size: 80%
Gradient Boosting Classifier	Number of estimators: 100 Learning rate: 0.1 Training set size: 80%

The classification performance is after that determined using metrics for precision, accuracy, F1-score, and recall, as shown in Table 3.

Table 3. Performance of CDDO-LEO across different classifiers

Classifier	Accuracy	Precision	Recall	F1-Score
Decision Tree	98.55%	98.66%	98.55%	98.54%
Random Forest	98.15%	98.09%	98.15%	97.96%
K-Nearest Neighbors (KNN)	90.30%	90.43%	90.30%	89.94%
Gradient Boosting	98.60%	98.89%	98.60%	98.55%

Of all the classifiers, Gradient Boosting had the best accuracy of 98.60% high grade performance and then Decision Tree with 98.55% which is close to adjacent Gradient boosting according to Figure 2. Random Forest classifier also exhibited a better performance of nearly 98.15% with KNN having a slightly less performance of 90.30%.

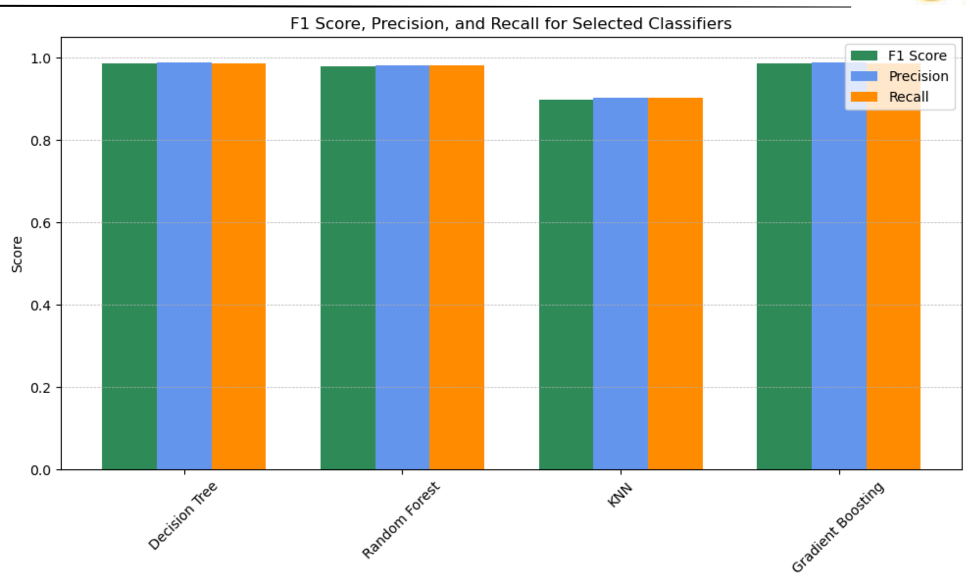

Figure 2. Performance matrices for different Classifier

These findings indicate that hybrid CDDO-LEO feature selection technique is useful in recalling the most informative features and lessening the computation complexity. The ability of the suggested hybrid CDDO-LEO of feature selection was comparatively evaluated against earlier methods of feature selection based on optimization.

In different studies, the hybrid of HGO and SSA with the use of GWO and WOA have been used to conduct feature selection for solving solve classification problems. The current procedures are ineffective to maintain the relatively high accuracy levels of the classification systems in case different systems are used.

Previous research has suggested divergence in the performance of metaheuristic feature selection techniques in various classifiers. As an example, a study of a hybrid HHO-SSA feature selection method revealed that standalone HHO produced a roughly 04% accuracy with KNN, 89% with SVM, and 94% with XGBoost, and the SSA and other optimizers (e.g., GWO, WOA) have variable accuracy across with same classifiers, highlighting sensitivity to classifier choice as well as dataset properties (e.g. accuracy differs between HHO and SSA and GWO and WOA with KNN/SVM/XGBoost) [24].

Also, the results of the HHOSSA-based studies revealed 96% and 98% accuracy with SVM and KNN and XGBoost, respectively, showing how different hybrid combinations of metaheuristics can change the results of the classifiers, as observed in Table 4 [25].

Table 4. Performance of Previously Used Feature Selection Methods

Classifier	HHO-SSA	GWO	WOA
KNN	94.64%	96.42%	94.64%
SVM	89.28%	82.14%	89.28%
XGBoost	94.64%	92.85%	96.42%

On the other hand, the proposed CDDO-LEO approach has expressed higher and more stable accuracy among the various classifiers especially when using ensemble learning models like Random Forest, Decision Tree, and Gradient Boosting as revealed in Table 5. In contrast to the past methods, CDDO-LEO is a more efficient way of using the informative features to reduce the dimensionality and improve the classification ability.

The best feature selection method of all previous methods is CDDO-LEO with the Gradient Boosting, and its accuracy was 98.60%. Moreover, the results of Decision Tree (98.55%)

and the random Forest (98.15%) improved significantly and surpass the traditional optimization-based methods.

Table 2. Performance of CDDO-LEO Feature Selection Method

Classifier	CDDO-LEO	CDDO- CDDO	LEO-LEO
KNN	90.30%	88%	88%
Random Forest	98.15%	98%	98%
Decision Tree	98.55%	96%	96%
Gradient Boosting	98.60%	95%	96%

Table 5 shows that compared to the previous techniques, CDDO-LEO is much more effective and reliable with ensemble-based classifiers, it is more accurate. As it is seen in the Fig. 1, the maximum accuracy of each method is given against the best-performing classifier.
As pointed out in the comparative study, the feature selection mechanism proposed by CDDO-LEO is better and more efficient in terms of feature subsets optimization than the HHO-SSA, GWO and WOA. Integrating exploration and refinement techniques used in CDDO and LEO, respectively, the proposed approach guarantees an improved performance of the classification with a more generalized process of feature selection, and is therefore the best option when it comes to intrusion detection and other high-dimensional classification problems.

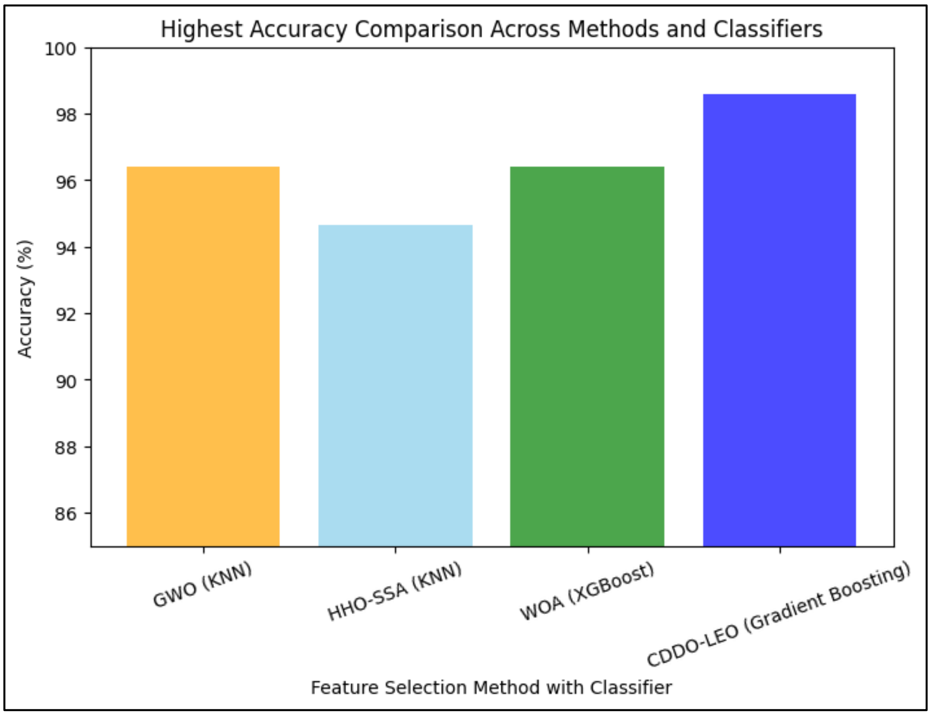

Figure 3. Comparing the Highest Accuracy Achieved by Each Method

As an additional measure of robustness of the suggested CDDO-LEO feature selection technique, a statistical significance of the classification results is provided. The model had a mean Accuracy of 0.9855 ± 0.0000 and mean F1-Score of 0.9854 ± 0.0000. Hypothesis testing is carried out to test whether the performance improvements observed are statistically significant.

The two obtained p-values are 0.000013 Accuracy and 0.000029 F1-Score, which are significantly less than the standard significance level of 0.05. These findings affirm that the improvements that have been made using the CDDO-LEO hybrid method are statistically significant and not due to random variation. Thus, it can be indicated that the proposed method is reliable and consistent in performing IoT intrusion detection tasks.
The suggested system will be based on an HP PROBOOK Laptop with the following specifications as illustrated in Table 6:

Table 3. System Specifications

Component	Specification
CPU	Intel(R) Core(TM) i5-7300U CPU @ 2.60GHz
RAM	16 GB
Operating System	Windows 10 (64-bit)
Programming Language	Python 10.13
IDE	Visual Studio Code

The system was tested on a hybrid algorithm (CDDO-LEO), which is a combination of Child Drawing Development Optimization algorithm (CDDO) and the Lagrange Elementary Optimization algorithm (LEO). Table 7 shows the processing time of the suggested hybrid algorithm and the standalone algorithms.

Table 4. Processing Time of Proposed Algorithms

Algorithm	Total Processing Time (Seconds)
CDDO	0.98
LEO	1.01
CDDO-LEO	2.02

The processing time of 2.02 seconds of the hybrid algorithm (CDDO-LEO) is slightly more than the standalone CDDO algorithm and the LEO algorithm. This processing time increment can be attributed to the integration of two optimization methods, which make the algorithm more robust and accurate. Although the hybrid method has a greater time to process, it has better classification performance and higher quality in its results than the standalone algorithms.

CONCLUSION

The proposed CDDO-LEO hybrid algorithm has a better feature selection because of the integration of the exploitation and exploration properties, leading to a high quality classification accuracy.
Experimental findings validate CDDO-LEO since it delivers superior performance than conventional methods, which reach 98.60% accuracy with Gradient Boosting as well as 98.55% with Decision Tree and 98.15% with Random Forest.
The statistical significance analysis supports that the suggested approach to the CDDO-LEO delivers the consistent and reliable improvements in the accuracy and F1-score, showing that the performance gains are not because of random variation. These findings demonstrate its reliable ability to select the appropriate features as well as a high level of computational capability.
Additionally, the comparative study shows better and steadier performance than HHO-SSA GWO and WOA in terms of maintaining accuracy across different classifiers. The hybrid algorithm demonstrates efficiency through processing time tests because it combines CDDO and LEO framework elements to deliver superior accuracy results without causing excessive computational load. The proposed approach will be expanded to new

domains that include biomedical data analytics and financial fraud detection and real-time intrusion systems through deep learning-based feature selection and parallel computing to boost performance speed and scalability.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

I would like to thank UoT, Al-Iraqia University for their support and contributions to this work.

Author Contributions

1. **MAA;** Conceptualization, Methodology, Writing – Original Draft.
2. **OHS**: Software Implementation, Experiments, Visualization.
3. **OSA:** Supervision, Writing – Review & Editing, Project Administration.
4. **SKM**: Supervision, Writing – Review & Editing, Project Administration.
5. **TAR**: Conceptualization, Project Administration, Supervision, Review & Editing

Ethical Approval

This article does not contain any studies involving human participants or animals performed by any of the authors.

Consent to Participate

Not applicable.

Data and Code Availability

The source code and datasets used in this study are publicly available in the following repository:
https://github.com/mustafaaabdulwahab/CDDO-Leo-Based-Feature-Selection-for-IoT-Intrusion-Detection

The repository contains the implementation of the proposed algorithm, dataset preprocessing scripts, and instructions for reproducing the experiments.

REFERENCES

[1] M. Injadat, A. Moubayed, A. B. Nassif, and A. Shami, "Multi-stage optimized machine learning framework for network intrusion detection," *IEEE Transactions on Network and Service Management*, vol. 18, no. 2, pp. 1803–1816, 2020.

[2] M. R. Aziz and A. S. Alfoudi, "Different mechanisms of machine learning and optimization algorithms utilized in intrusion detection systems," in *AIP Conference Proceedings*, vol. 2839, no. 1, Sep. 2023.

[3] M. Conti, A. Dehghantanha, K. Franke, and S. Watson, "Internet of Things security and forensics: Challenges and opportunities," *Future Generation Computer Systems*, vol. 78, pp. 544–546, 2018.

[4] A. M. Aladdin and T. A. Rashid, "LEO: Lagrange elementary optimization," *arXiv preprint arXiv:2304.05346*, 2023.

[5] A. A. Ameen, T. A. Rashid, and S. Askar, "A tutorial on child drawing development optimization," pp. 153–167, 2023, doi: 10.2991/978-94-6463-110-4_12.

[6] A. F. Ali and M. A. Tawhid, "A hybrid PSO and DE algorithm for solving engineering optimization problems," *Applied Mathematics & Information Sciences*, vol. 10, no. 2, pp. 431–449, 2016, doi: 10.18576/amis/100207.

[7] S. Cho, S. Lim, C. Kim, S. Wi, T. Kwon, W. S. Youn, and S. Cho, "Enhancement of soft-tissue contrast in cone-beam CT using an anti-scatter grid with a sparse sampling approach," *Physica Medica*, vol. 70, pp. 1–9, 2020.
[8] M. Allam and M. Nandhini, "Optimal feature selection using binary teaching learning based optimization algorithm," *Journal of King Saud University–Computer and Information Sciences*, vol. 34, no. 2, pp. 329–341, 2022.
[9] B. M. Rashed and N. Popescu, "Machine learning techniques for medical image processing," in *Proc. 2021 International Conference on e-Health and Bioengineering (EHB)*, 2021, pp. 1–4.
[10] M. Abdel-Basset, W. Ding, and D. El-Shahat, "A hybrid Harris Hawks optimization algorithm with simulated annealing for feature selection," *Artificial Intelligence Review*, vol. 54, no. 1, pp. 593–637, 2021.
[11] M. Tubishat, M. Alswaitti, S. Mirjalili, M. A. Al-Garadi, M. T. Alrashdan, and T. A. Rana, "Dynamic butterfly optimization algorithm for feature selection," *IEEE Access*, vol. 8, pp. 194303–194314, 2020, doi: 10.1109/ACCESS.2020.3033757.
[12] H. R. R. Zaman and F. S. Gharehchopogh, "An improved particle swarm optimization with backtracking search optimization algorithm for solving continuous optimization problems," *Engineering with Computers*, vol. 38, suppl. 4, pp. 2797–2831, 2022.
[13] P. G. Luan and N. T. Thinh, "Hybrid genetic algorithm based smooth global-path planning for a mobile robot," *Mechanics Based Design of Structures and Machines*, vol. 51, no. 3, pp. 1758–1774, 2023.
[14] J. A. Widians, R. Wardoyo, and S. Hartati, "A study on text feature selection using ant colony and grey wolf optimization," in *Proc. 2022 Seventh International Conference on Informatics and Computing (ICIC)*, 2022, pp. 1–7.
[15] V. Gupta and E. Kumar, "H3O-LGBM: Hybrid Harris hawk optimization based light gradient boosting machine model for real-time trading," *Artificial Intelligence Review*, vol. 56, no. 8, pp. 8697–8720, 2023.
[16] F. W. Liu and C. Hu, "Research on swarm intelligence optimization algorithm," *Journal of China Universities of Posts and Telecommunications*, vol. 27, no. 3, p. 1, 2020.
[17] D. A. Kristiyanti, I. S. Sitanggang, and S. Nurdiati, "Feature selection using new version of V-shaped transfer function for salp swarm algorithm in sentiment analysis," *Computation*, vol. 11, no. 3, Art. no. 56, 2023.
[18] O. Gilo, J. Mathew, S. Mondal, and R. K. Sanodiya, "Unsupervised sub-domain adaptation using optimal transport," *Journal of Visual Communication and Image Representation*, vol. 94, Art. no. 103857, 2023.
[19] K. K. Ghosh, R. Guha, S. Ghosh, S. K. Bera, and R. Sarkar, "Atom search optimization with simulated annealing—A hybrid metaheuristic approach for feature selection," *arXiv preprint arXiv:2005.08642*, 2020.
[20] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," *Future Generation Computer Systems*, vol. 97, pp. 849–872, 2019.
[21] M. A. Al-Betar, A. K. Abasi, G. Al-Naymat, K. Arshad, and S. N. Makhadmeh, "Bare-bones based salp swarm algorithm for text document clustering," *IEEE Access*, vol. 11, pp. 100010–100028, 2023.
[22] Q. Al-Tashi, H. Md Rais, S. J. Abdulkadir, S. Mirjalili, and H. Alhussian, "A review of grey wolf optimizer-based feature selection methods for classification," in *Evolutionary Machine Learning Techniques: Algorithms and Applications*, 2020, pp. 273–286.

[23] E. C. P. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu, and A. A. Ghorbani, "CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment," *Sensors*, vol. 23, no. 13, Art. no. 5941, 2023. [Online]. Available: https://www.kaggle.com/datasets/subhajournal/iotintrusion
[24] L. Wen and Y. Cao, "A hybrid intelligent predicting model for exploring household CO2 emissions mitigation strategies derived from butterfly optimization algorithm," *Science of the Total Environment*, vol. 727, Art. no. 138572, 2020, doi: 10.1016/j.scitotenv.2020.138572.
[25] S. Arora, H. Singh, M. Sharma, S. Sharma, and P. Anand, "A new hybrid algorithm based on grey wolf optimization and crow search algorithm for unconstrained function optimization and feature selection," *IEEE Access*, vol. 7, pp. 26343–26361, 2019, doi: 10.1109/ACCESS.2019.2897325.
[26] A. Moslemi and A. Ahmadian, "Subspace learning for feature selection via rank revealing QR factorization: Unsupervised and hybrid approaches with non-negative matrix factorization and evolutionary algorithm," *arXiv preprint arXiv:2210.00418*, 2022.