ISSN: 1803-3814 (Printed), 2571-3701 (Online)
https://doi.org/10.13164/mendel.2023.1.062

A Study on Heuristic Algorithms Combined With LR on a DNN-Based IDS Model to Detect IoT Attacks

Thanh - Thuy Tran Thi[1], Duc - Thuan Luong[2], Hong - Duc Nguyen[2], and Trong - Minh Hoang[2,✉]

[1]Faculty of Electronic Engineering 1, Posts and Telecommunication Institute of Technology, Hanoi, Vietnam
[2]Faculty of Telecommunication 1, Posts and Telecommunication Institute of Technology, Hanoi, Vietnam
hoangtrongminh@ptit.edu.vn✉, thuyttt@ptit.edu.vn, thuanld.b19vt399@stu.ptit.edu.vn, ducnh.b19vt096@stu.ptit.edu.vn

Abstract

Current security challenges are made more difficult by the complexity and difficulty of spotting cyberattacks due to the Internet of Things explosive growth in connected devices and apps. Therefore, various sophisticated attack detection techniques have been created to address these issues in recent years. Due to their effectiveness and scalability, machine learning-based Intrusion Detection Systems (IDSs) have increased. However, several factors, such as the characteristics of the training dataset and the training model, affect how well these AI-based systems identify attacks. In particular, the heuristic algorithms (Genetic Algorithm-GA, Particle Swarm Optimization-PSO, Cuckoo Search Optimization-CSO, Firefly Algorithm-FA) optimized by the Logistic Regression (LR) approach employ it to pick critical features of a dataset and deal with data imbalance problems. This study offers an Intrusion Detection System (IDS) based on a deep neural network and heuristic algorithms combined with LR to boost the accuracy of attack detections. Our proposed model has a high attack detection rate of up to 99% when testing on the IoT-23 dataset.

Keywords: *Intrusion Detection System, Deep Neuron Network, Heuristic Algorithms, IoT-23 dataset.*

Received: 04 June 2023
Accepted: 16 June 2023
Online: 21 June 2023
Published: 30 June 2023

1 Introduction

The Internet of Things (IoT) technologies have linked billions of items and amassed a vast amount of data, which may be utilized for automated processes and intelligent computing to integrate the physical and digital worlds. These days, IoT applications impact practically every element of society, including industry, healthcare, transportation, and agriculture [11, 26]. With its benefits, new difficulties are presented to network security and performance [24].

An intrusion detection system (IDS) can identify unusual network activities brought on by successful assaults as the initial security measure. Because of the variety of devices, the ease of usage of the protocols, and the constrained resources of the system components, IoT systems might benefit from IDS [8]. IDS systems with machine learning-based active forms use past data from datasets to compare and identify real threats [25]. Accumulated datasets greatly influence the accuracy of predictions and their application in real-life scenarios and the intelligence of machine learning algorithms. Recently proposed neural network-based IDS systems use KDD-99, NSL-KDD, and UNSW-NB15 datasets to assess and identify a variety of threats on computer and IoT networks [24, 30, 6].

Deep learning is a type of machine learning that enables computers to understand the world using a hierarchy of ideas and learn from experience. Deep neural networks (DNN) are networks created using deep learning. Deep learning applications for issues like prediction, identification, and optimization have shown precise results and stimulated a sizable amount of research using this approach [27]. IDS systems utilizing DNN have gotten much interest in recent research [28]. These suggested systems are frequently paired with data preparation techniques to increase performance. However, the accuracy and scope of the application of DNN-based IDS systems have long been a problem for researchers.

This study suggests using an LR-heuristic (GA, PSO, CS, FA) technique and a DNN-based IDS system to improve the attack detection rate in an IoT context. Before DNN training, our proposed model's a heuristic technique is optimized using Logistic Regression (LR) that chooses the most important characteristics from a dataset. On the IoT-23 [20] datasets, the suggested model has been evaluated, and in particular, our method has solved the major obstacle of the accuracy score of the DNN-based IDS system. When using data balancing techniques, the test accuracy on the IoT-23 dataset is 99%.

2 Related Work

In recent years, IDS-based machine learning systems to prevent DDOS attacks have experienced significant growth due to their effectiveness in facing various attacks from a vast number of IoT devices. The authors of [23] proposed an IDS based on Deep Learning (DL) using Feed Forward Deep Neural Networks (FFDNN) and a filter-based feature selection approach. Using the well-known NSL-KDD dataset, the FFDNN-IDS was evaluated and compared to existing machine learning techniques: Support vector machines (SVM), Decision Tree (DT), K - Nearest Neighbor (KNN), and Naive Bayes (NB). Test results indicate that FFDNN-IDS is more accurate than other techniques. Using the NSL-KDD dataset, an accelerated DNN architecture was developed [21] to detect anomalies in network data. It demonstrates that DNN-based IDS can reliably identify certain attack classes (DoS and Probe assaults) with the required training samples but cannot effectively classify attack types (R2L and U2R) with limited training samples. In [5], a DNN model for intrusion detection is proposed, along with a novel preprocessing technique tested on the KDDCup'99 dataset and designed to improve the performance of detection algorithms. The test results indicate that their pretreatment method outperforms conventional preprocessing techniques regarding precision, recall, and F1-score, resulting in more accurate detection of Advanced Persistent Threats (APTs).

The authors of [7] classify deep learning models used for intrusion detection and summarize the relevant academic literature. Using two historical datasets (KDD 99 and NSL-KDD) and two current datasets, the authors trained and evaluated four primary deep learning models for the classification task type: feedforward neural networks, autoencoders, deep-belief networks, and short-term, long-term memory networks (CIC-IDS2017, CIC-IDS2018). In all four datasets, their findings demonstrated that deep feedforward neural networks generate the appropriate evaluation metrics for accuracy, F1-scores, and training and inference times. The results also demonstrated that the supervised feedforward neural network does not perform better than two well-known semi-supervised learning models, the autoencoder and the deep belief network. By selecting potential features before network input processing, [14] aims to enhance the performance of Deep Neural Networks (DNNs). This work employs the KDD Cup 99 dataset, one of the standard datasets for intrusion detection. Based on their experiments' results, they concluded that feature selection improves IDS compared to the technique without feature selection. This research demonstrated that the DNN for IDS might improve accuracy by 99.4%, precision by 99.7%, recall by 97.9%, and F1-score by 98.8%. However, the diversity and updating of datasets in IoT today require researchers to continue looking for and developing new attack-prevention models.

One of the critical performance objectives of an ML-based IDS system is to improve attack detection performance parameters and reduce computation time complexity. Therefore, machine learning application schemes in IDS systems are often combined with other techniques. Specifically, the swarm approximation optimization techniques. The authors of [9] proposed an IDS model that used an ANN model with LR-GA to achieve a work detection rate of up to 95.26%. The ANN parameters in this model are optimized using the GSPSO technique. The authors in [12] identified the key characteristics of the data and assisted in modeling it by using GA to determine the best parameters for the decision tree (DT) and K-nearest neighbor (KNN) algorithms. They can identify DDOS attacks up to 99% of the time. However, other sorts of attacks are now far less accurately detected. In [2], the GA method is used with a support vector machine (SVM) to choose characteristics from 10 categories with three different priority levels. The authors of [32] described a model that uses a more advanced GA algorithm and deep belief networks (DBN). They assessed the outcomes using the NSL-KDD dataset, demonstrating how their methodology may simplify the network architecture and increase attack detection accuracy. Their study suggested and presented three experiments that combined genetic algorithm and logistic regression. Their suggested model has an F1-score of 93.56% and a classification accuracy of 94.55%.

Synthetic minority sampling technique (SMOTE) and particle swarm optimization (PSO) are combined in the suggested classification algorithms in [29], which also incorporate many well-known classifier techniques, including logistic regression, decision tree model C5 (C5), and 1-nearest neighbor. In order to address the randomly occurring problem in a cloud computing context and improve Initial Aerofoil Geometry, the authors in [19, 18] reviewed the available modified PSO scheduling methods. The comparison demonstrates that the new algorithm successfully improves upon the original PSO. The Bayesian information criterion (BIC) was suggested in [22] with the logistic regression model and the particle swarm optimization algorithm. Investigation and comparison of the effectiveness of various fitness functions using BIC. Several different kinds of datasets are used to assess the performance of the suggested technique. Results from experiments utilizing various dataset types show how our suggested strategy can dramatically enhance classification performance with few features. The outcomes demonstrate that the suggested approaches competitively outperform other current fitness functions. [15] studies and develops associated technical levels and realizes early diagnosis and intervention of illness risk factors through applying a unique quantum particle swarm optimization. The findings of the comparative experiments confirm that the proposed strategy is suitable for creating a Logistic Regression Health Assessment Model.

The weighted binary cuckoo search (WBCS) algo-

rithm and logistic regression algorithm were used to produce the best results for the speech emotion recognition (SER) test, according to the authors of [33]. In order to ensure that the characteristics adapt to various circumstances, cross-training is used. The effective classification scores are good for less than 100 training sets. The suggested approach is more accurate than current intelligent optimization dimensionality reduction algorithms. With an accuracy of up to 90.30%, the experimental results in [17] demonstrate that the cuckoo search algorithm in the FNN model can be used to forecast possible financial troubles.

The Penalized Support Vector Machine (PSVM) tuning parameter will be found using a new hybrid firefly algorithm and particle swarm optimization [1]. In order to discover the most appropriate descriptors with the best classification performance, their suggested technique may effectively use the advantages of both firefly and particle swarm algorithms. The test results on four benchmark QSAR datasets demonstrate the proposed algorithm's improved classification accuracy and the number of descriptors selected compared to competing approaches. [31] establishes a comprehensive FA-SVM supply chain financial credit risk assessment model. The following are the key points of innovation: When FA-SVM picks the input variables in the algorithm setting, the process is random, improving the model's randomness without following the dataset's initial order.

The survey results above have shown that machine learning algorithms or heuristic algorithms [10] achieved specific achievements in some fields, including cybersecurity in IoT. However, the proposals are aimed at the dataset of computer networks in general, not IoT systems in particular. In addition, the dataset they use is relatively old, not diverse, and has not been updated with new attack types or features of the current IoT system. Moreover, machine learning or heuristic algorithms are implemented individually, so maximum efficiency has not been achieved. Therefore, this study will address the above vulnerability to evaluate the implementation of heuristic algorithms combined with LR on the DNN-based IDS model for the IoT-23 dataset.

3 Background

The Logistic Regression (LR) algorithm constitutes a supervised learning approach that functions to predict the output of a dependent variable using independent variables and their relationships. LR model trains on labeled data and then estimates the coefficients of independent variables to generate a classification model. The LR algorithm owed its roots to the 19th century when statisticians formulated regression methods for data analysis. In the early versions of the LR, categorizing statistical problems was vital; however, in the 1940s and 1950s, the method was further refined by mathematicians and statisticians, including Joseph Berkson, David Cox, and William Cochran. In recent years, LR gained immense popularity in medical applications, predicting the probability of dependent variables such as the possibility of getting an infection or developing heart disease. In business and financial applications, such as predicting the likelihood of a customer buying a product, assessing the risk of investing, and forecasting the cost of developing a new product, LR became a required statistical method. LR is a widely used algorithm in machine learning, mainly in binary classification problems, such as email classification, credit fraud detection, and product categorization prediction, making it one of the foundational algorithms in machine learning. In addition, LR also handles the problem of data imbalance [16].

GA, PSO, CS, and FA algorithms have been widely applied in various fields followed benchmarking standards [13], including function optimization, combinatorial optimization, artificial neural networks, and other optimization problems. The Genetic Algorithm (GA) is a global search algorithm based on the principles of natural evolution, developed based on Charles Darwin's theory of evolution. GA is used to optimize objective functions that do not have a mathematical analysis solution or are too complex to compute. The GA is a powerful global search method with limitations, such as slow convergence for significant problems and the possibility of falling into the local optimum.

The Particle Swarm Optimization (PSO) algorithm is an optimization method inspired by the movement of a flock of birds searching for food in space. The PSO algorithm is a powerful global search method capable of quickly searching in large search spaces. This algorithm uses "particles" to find and optimize solutions. However, it also has limitations, such as the ability to fall into the local optimum and slow convergence for large-sized problems.

The Cuckoo Search (CS) algorithm [3] is a global search algorithm based on the breeding behavior of pigeons. It was inspired by how pigeons put their eggs in other birds' nests to lay and deceive host birds. The Cuckoo Search algorithm can find the best solution in ample search space, especially in intermittent and non-derivative optimization problems. However, it also has some limitations, such as the slow convergence speed for large-sized problems and the need for a long time to find the optimal solution.

The Firefly Algorithm [4] is an optimization method inspired by the way fireflies light up in the dark to find food and communicate with each other. This algorithm uses "fireflies" to find and optimize solutions.

4 Proposal Model

The main components of the proposed model include the K-mean method, Logistic regression (LR) algorithm, heuristic algorithms, and DNN, illustrated in Figure 1. Attack types are separated into high-traffic attacks and low-traffic assaults during the data preparation phase using the provided data set IoT-23. Data set partitioning employs processing methods that reduce computing complexity and data area size. The

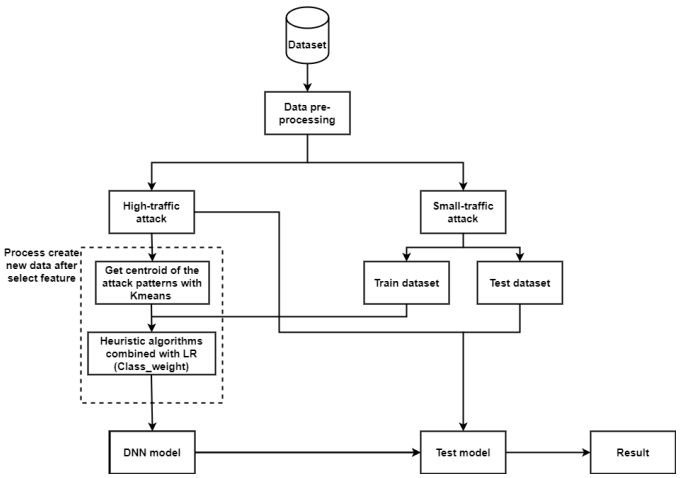

Figure 1: The proposal model

Table 1: IoT-23 dataset.

Number	Type	Samples
0	Attack	9398
1	Benign	30854315
2	C&C	21995
3	C&C FileDownload	53
4	C&C HeartBeat	33673
5	C&C HeartBeat Attack	834
6	C&C HeartBeat FileDownload	11
7	C&C Mirai	2
8	C&C PartOfAHorizontalPortScan	888
9	C&C Torri	30
10	DdoS	19538713
11	FileDownload	18
12	Okiru	60990708
13	Okiru Attack	3
14	PartOfAHorizontalPortScan	213852924
15	PartOfAHorizontalPortScan Attack	5
Sum		325303570

data set is split 80/20 into training and test data for assaults with low traffic. For this study, we employ the standard clustering method, reducing the dataset's dimensionality for high-traffic attacks. When clustering the data, we use the K-means technique to lessen the amount of duplicated data while still covering the complete data domain. We use the LR paired with heuristic algorithms to optimize the heuristic algorithm's parameters and reduce the data dimension. The target data is then prepared for categorization by the DNN model by combining low-traffic and high-traffic assault data. At this point, data balancing processes are also used on the training set of the proposed model to remove distinctive biases, and once the data has been processed, the suggested DNN model will be evaluated. The in-depth processing methods are presented in detail in this part 4.

4.1 Data Preprocessing

The IoT-23 dataset is a good network traffic dataset from Internet of Things (IoT) devices published in 2020. The collection of datasets excludes local network information and instead focuses on network properties to facilitate generalization to all IoT networks. The IoT-23 database functions as a valuable resource for researchers and practitioners in the field of IoT security by providing a comprehensive dataset for testing and evaluating security solutions, including the detection and prevention of various types of IoT attacks. It has 20 malware captures executed in IoT devices and three captures for benign IoT device traffic. The dataset includes 23 features with 16 attack types described in Table 1, of which 15 are malicious layers and one is safe layer.

When conducting data processing, the missing data is processed by getting the average values of each attack type related to its features. As a result, the mean corresponding to the 16 attack types will be used to fill in the missing values in the feature column. Some features, "unnamed", "label," "local resp," and "local orig" in the IoT-23 dataset, are eliminated from the read records because they are null and have no significance.

Types, connection types, services, and other network properties are analyzed and encoded for translation to numeric data. Categorical data is made up of columns like "service," "proto," and "conn state," each of which transforms the various values it contains into a separate, binary state vector [32]. Therefore, the number of features can be increased by this encoding. The data file presents two columns of the IPv4 address type, "id.orig h" and "id.resp h," each representing the value of a separate IP address. The IP address encoding standard encodes IPv4 address data into a digital format using the 'ipaddress' library. By converting each octet to an 8-bit binary and then concatenating four octets to create a 32-bit binary string, this library will convert IP addresses to a digital representation. Then we will convert back to decimal from 32-bit binary. The missing data is processed by getting the average values of each attack type related to its features. As a result, the mean corresponding to the 16 attack types will be used to fill in the missing values in the feature column. In this paper, we extract four types of attacks from the IoT-23 dataset displayed in Table 2. We have extracted two popular attack types with high traffic and two with low traffic, Okiru, PartOfAHorizontal-PortScan, C&C PartOfAHorizontal- PortScan, C&C-Heart Beat Attack, labeled from 0 to 3 to test with the proposed model.

4.2 Data Clustering

In this scheme, we have clustered the dataset after preprocessing the data. For high-traffic attacks, it is essential to find important data points that have much

Table 2: The dataset used in the proposed model is extracted from the IoT-23 dataset.

Number	Type	Samples
0	Okiru	100000
1	PartOfAHorizontalPortScan	100000
2	C&C PartOfAHorizontal-PortScan	888
3	C&C HeartBeat Attack	834
Sum		201722

influence on the model instead of dealing with the entire dataset. Thus, clustering high-traffic datasets reduces the size of the dataset and selects the critical data points. In this study, we use the K-means method for high-traffic attacks. The algorithm will work as follows, initially choosing K random cluster centers. It then computes the data points for those original cluster centers if the data points closest to the cluster center will belong to that cluster. The cluster center is recalculated with the mean of the data points. The algorithm stops when it cannot be improved further. The steps of the K-means algorithm work as algorithm 1.

Algorithm 1 K-means algorithm

1: Initialize k initial center points. These center points are chosen randomly or based on basic knowledge of the data
2: Calculate each data point's distance to the center point, then assign it to the group with the closest central point
3: Recalculate the position of k-center points based on the data points assigned to the group
4: Repeat steps 2 and 3 until the center points are unchanged or the maximum number of iterations previously given is reached
5: Returns groups of data that have been classified

4.3 Features Selection

In the IoT-23 dataset, the number of features is 33 after data preprocessing. In the fact that the IoT-23 dataset contains small traffic attacks that lead to data imbalances, making the neural network challenging to detect and confusing between types of attacks. Therefore, our study uses LR combined with various heuristic techniques to choose a small practical set of features and reduce dimensionality to enhance attack detection accuracy on the IoT-23. Moreover, it will also deal with the data imbalance problem using class weight methods.

The GA searches for potential solutions and uses crossovers and mutations to generate new generations of solutions, in which better solutions are preferred for the next generation. Solutions are represented as gene sequences, similar to the characters in a string. The LR-GA algorithm is an algorithm that combines the LR algorithm into the fitness function of GA. The LR-GA algorithm has the main steps in algorithm 2.

Algorithm 2 Optimize Genetic Algorithm parameters by LR

1: Initialize a set of initial parameters
2: Initialize the instance for the initial population
3: Calculation of fitness function based on LR algorithm
4: Select highly effective individuals
5: Apply mutation to some solutions in the new generation to create diversity in the solution set
6: Evaluate the quality of each solution in the new generation
7: Select the best solutions to become the solution set for the next generation
8: Repeat from step 3 to step 6 until the best solution or a sufficient number of generations is reached
9: End

The LR-PSO algorithm is an algorithm that combines the LR algorithm into the fitness function of PSO. The LR-PSO algorithm has the main steps of algorithm 3.

Algorithm 3 Optimize Particle Swarm Optimization parameters by LR

1: Initialize a set of initial parameters
2: Initialize the instance for the initial population
3: Calculate the fitness function based on the LR algorithm and choose the best position of the individuals as the next population
4: Update the position of each particle by computing a speed vector and adding it to the particle's current position. The speed vector is calculated by combining two components: (a) the free component, where the particle is guided by the best position it has found, and (b) the social component, where the particle is guided by the best position the other particles in the set have found
5: Evaluate the quality of each grain after updating the position
6: Update the best position that the set of particles has found
7: Repeat step 3 to step 6 until the best solution is reached or a sufficient number of iterations are reached
8: End

The Cuckoo Search algorithm combined with the Logistic Regression algorithm is presented in algorithm 4.

The Firefly algorithm combined with the Logistic Regression algorithm is implemented as algorithm 5.

4.4 Deep Neural Network-based IDS Model

A DNN model with five layers—one input layer, three hidden layers, and one output layer—is constructed in this study. Initially, the layer will have 128 nodes, which will be compressed to 64, 32, and 16 nodes. The output will then depend on the number of classes in the processing dataset. This neural network's nonlinear function is called "Relu". To improve the learning model, we employ the Dropout approach to randomly turn off some nodes during training at a rate of 5% after each lesson. In the network, the output layer will have four nodes corresponding to the problem's output number. We use the term "softmax" for the activation

function of this last layer. The "Adam" optimizer and the cross entropy function simulate and put the optimization function into practice. The learning rate is vital for the learning model; we initially set it to 0.001 to allow it to learn; if the model deteriorates or validation errors rise for three consecutive periods, we drop the learning rate by 0.2 times to a minimum of 0.00001. We employed a method that prompted the suggested model to stop training early to reduce over-learning. The training is over when the validation error function does not decrease after a certain number of epochs. To verify that the validation error does not increase over the epochs, we can adjust the number of epochs to provide the best network accuracy over the learning interval. Lastly, using the test IoT-23 dataset, we will assess the proposed model. Loss, precision, recall, and F1-score are the criteria that are examined.

Algorithm 4 Optimize Cuckoo Search parameters by LR

1: Initialize a cuckoo-searchable pair of pants that includes random values for the parameters of the Logistic Regression model
2: Evaluate each bird's clothing performance using the Logistic Regression model to predict and compare with an actual value
3: Use the breeding strategy of the cuckoo search to create the next generation of populations by combining the values of the birds with the best effect
4: Repeat steps two and step 3 until the best stopping criterion is reached
5: End

Algorithm 5 Optimize Firefly Algorithm parameters by LR

1: Initialize Fireflies. Initialize a large number of initial fireflies. Set parameters, including the number of fireflies, alpha parameter (similar arithmetic with coefficients in logistic regression algorithm), distance r
2: Search and move. Calculate the objective function value (usually the cost function) for each firefly based on the parameters of the logistic regression model
3: Evaluation and selection. Performance evaluation of logistic regression model with searched parameters. Please select the best fireflies based on their objective function value
4: Update the model. Use the parameters of the best fireflies to update the logistic regression model
5: Repeat steps 2 to 4 until the performance requirements of the model are met
6: End

5 Numerical Results and Discussion

Table 3 shows the parameters that we use for this proposed model. In this study, we use the features selection technique to reduce the dimensionality of the dataset, which leads to choosing a small practical set of features, to increase the speed without significantly reducing the accuracy. Following are the models that we have tested on our little IoT-23 set. The training models have been opened with 'class_weight' weights, which helps the models to work effectively with imbalanced datasets. We use the numerical parameters to evaluate the proposed method: precision, recall, and f1-score.

Table 3: Model parameters.

Hyperparameters	Value
Number of generations	10
Population size	50
Balance mode	1

The performance of the proposed model is shown in Table 4. The results of the four heuristic algorithms used are generally relatively good. However, of the four algorithms for feature selection above, the LR-GA algorithm gives the best results. Because after features selection, the LR-GA algorithm selects only eight features from the original 33 features, still gives high-performance parameters up to 99%, and reduces computation time. The LR-GA algorithm will reduce the computation time on the test set because it only has to decide based on eight features, but the training process needs more time because it has to choose eight features from the original 33 features. In other algorithms, the remaining features are 30, 26, and 26 for LR-PSO, LR-CSO, and LR-FAO algorithms.

The training process of the four models shown in Figures 2, 3, 4, and 5 shows that the training results of the algorithms are relatively good.

Finally, Table 5 displays the experimental performance outcomes (precision, recall, F1 score) of each class of each algorithm for the proposed model that was verified on the IoT-23 dataset. We can see that the proposed model has outstanding metrics when used on small samples. As a result, it is preferred to be applied to devices with low resources at edge networks.

Table 4: The performance of the proposed model on the IoT-23 dataset.

Method	Number of remaining features	Precision	Recall	F1-score
LR-GA	8	0.99	0.99	0.99
LR-PSO	30	0.99	0.99	0.99
LR-CSO	26	0.99	0.99	0.99
LR-FAO	26	0.99	0.99	0.98

This paper proposes a DNN-based intrusion detection system (IDS) model with heuristic methods (GA, PSO, CSO, and FAO) combined with LR to detect and evaluate attacks. IThe proposed model uses the K-means clustering method to reduce the data size. Heuristic algorithm's parameters are optimized by LR to extract important features while using class-weight techniques to prevent imbalance attacks. Attacks are

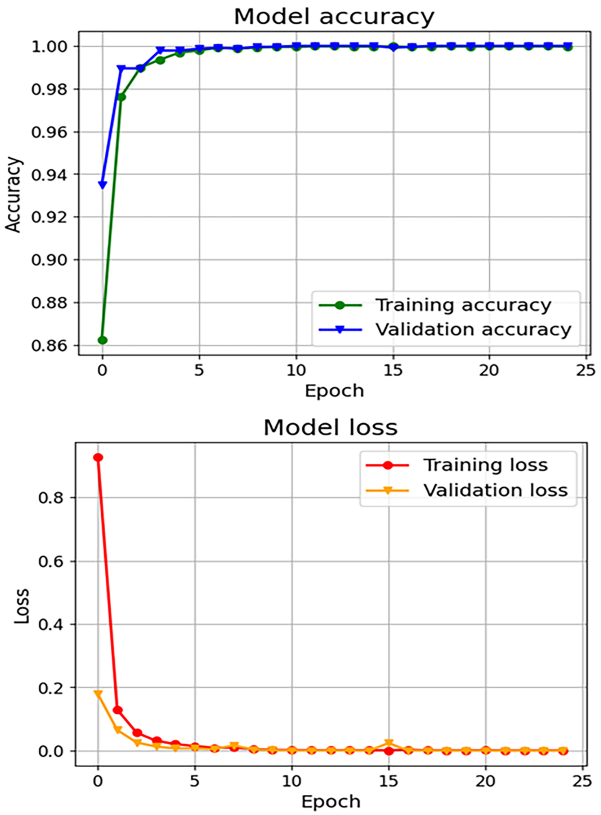

Figure 2: Training graph of LR-GA method (including accuracy and loss)

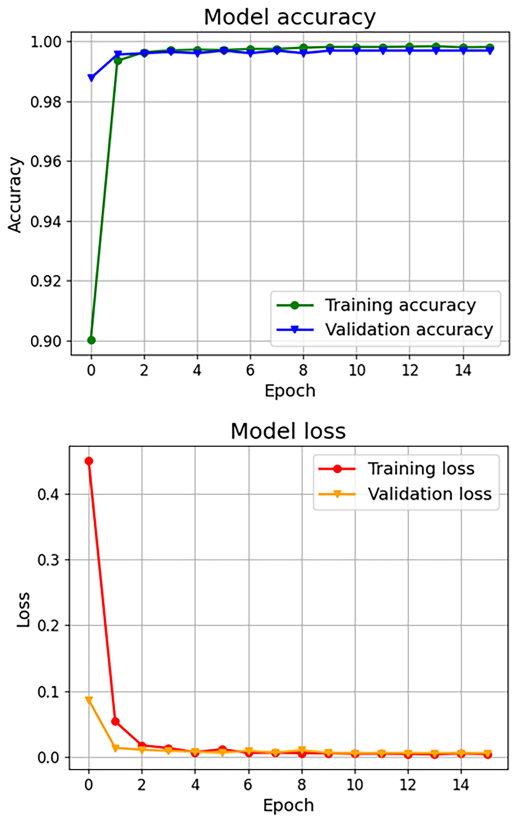

Figure 3: Training graph of LR-PSO method (including accuracy and loss)

Table 5: The performance of the proposed model on the IoT-23 dataset by class.

Method	Class	Precision	Recall	F1-score	Smaples
LR-GA	0	0.99	0.99	0.99	100000
	1	0.99	0.99	0.99	100000
	2	0.99	0.98	0.99	177
	3	0.98	0.99	0.99	168
LR-PSO	0	0.99	1	0.99	100000
	1	1	0.99	0.99	100000
	2	0.99	1	1	177
	3	0.99	0.99	0.99	168
LR-CSO	0	1	0.99	0.99	100000
	1	0.99	0.99	0.99	100000
	2	1	0.99	1	177
	3	0.99	1	1	168
LR-FAO	0	1	0.99	0.99	100000
	1	0.99	0.98	0.99	100000
	2	0.99	0.99	0.99	177
	3	1	0.99	0.99	168

differentiated based on DNN. On the IoT-23 dataset, the proposed model has been evaluated using a variety of sample sizes. The test results demonstrate that, for all heuristic methods, our proposed model's attack de-

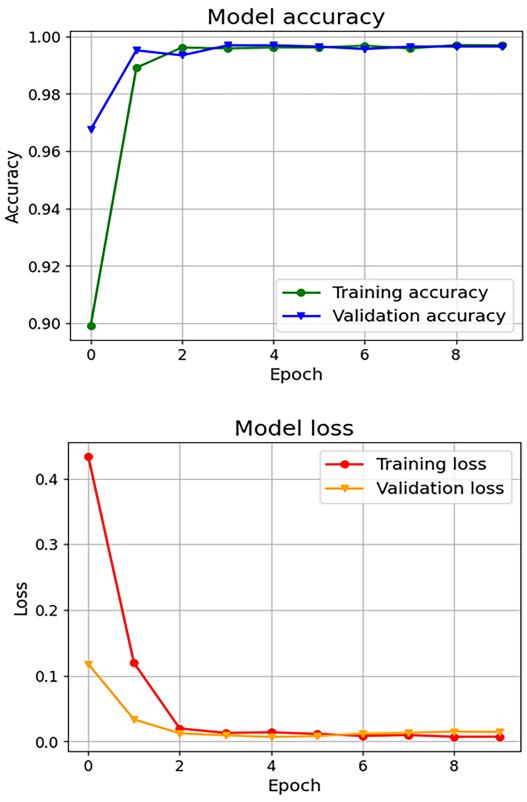

Figure 4: Training graph of LR-CSO method (including accuracy and loss)

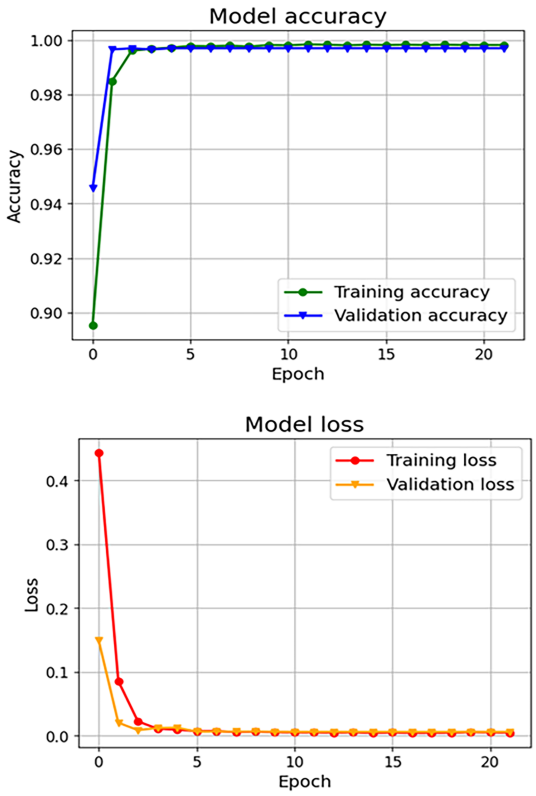

Figure 5: Training graph of LR-FA method (including accuracy and loss)

tection accuracy up to 99.9% with a modest number of samples.

Moreover, the GA algorithm combined with LR on a DNN-based IDS model for attack detection is the most effective of the heuristic algorithms beacause the number of features selected is relatively small (8 features) but still ensures accuracy, thereby reducing the computation time. Consequently, the proposed model applies to network border devices with limited resource availability, particularly the model employing the GA algorithm. The following steps will entail the deployment of real-time devices in agricultural IoT systems.

Acknowledgement: VINIF partly supports this work. Tran Thi Thanh Thuy was funded by the Master, Ph.D. Scholarship Programme of Vingroup Innovation Foundation (VINIF), code VINIF.2022.ThS.087.

References

[1] Al-Thanoon, N. A., Qasim, O. S., and Algamal, Z. Y. A new hybrid firefly algorithm and particle swarm optimization for tuning parameter estimation in penalized support vector machine with application in chemometrics. *Chemometrics and Intelligent Laboratory Systems 184* (2019), 142–152.

[2] Alshamy, R., Ghurab, M., Othman, S., and Alshami, F. Intrusion detection model for imbalanced dataset using smote and random forest algorithm. In *Advances in Cyber Security: Third International Conference, ACeS 2021, Penang, Malaysia, August 24–25, 2021, Revised Selected Papers 3* (2021), Springer, pp. 361–378.

[3] Camacho-Villalón, C. L., Dorigo, M., and Stützle, T. An analysis of why cuckoo search does not bring any novel ideas to optimization. *Computers & Operations Research 142* (2022), 105747.

[4] Camacho-Villalón, C. L., Dorigo, M., and Stützle, T. Exposing the grey wolf, moth-flame, whale, firefly, bat, and antlion algorithms: six misleading optimization techniques inspired by bestial metaphors. *International Transactions in Operational Research* (2022).

[5] Chen, P., Guo, Y., Zhang, J., Wang, Y., and Hu, H. A novel preprocessing methodology for dnn-based intrusion detection. In *2020 IEEE 6th International Conference on Computer and Communications (ICCC)* (2020), IEEE, pp. 2059–2064.

[6] Choudhary, S., and Kesswani, N. Analysis of kdd-cup'99, nsl-kdd and unsw-nb15 datasets using deep learning in iot. *Procedia Computer Science 167* (2020), 1561–1573.

[7] Gamage, S., and Samarabandu, J. Deep learning methods in network intrusion detection: A survey and an objective comparison. *Journal of Network and Computer Applications 169* (2020), 102767.

[8] Heidari, A., and Jabraeil Jamali, M. A. Internet of things intrusion detection systems: A comprehensive review and future directions. *Cluster Computing* (2022), 1–28.

[9] Hosseini, S. A new machine learning method consisting of ga-lr and ann for attack detection. *Wireless Networks 26* (2020), 4149–4162.

[10] Kanagaraj, G., Masthan, S. S., and Vincent, F. Y. Meta-heuristics based inverse kinematics of robot manipulator's path tracking capability under joint limits. In *Mendel* (2022), vol. 28, pp. 41–54.

[11] Kashani, M. H., Madanipour, M., Nikravan, M., Asghari, P., and Mahdipour, E. A systematic review of iot in healthcare: Applications, techniques, and trends. *Journal of Network and Computer Applications 192* (2021), 103164.

[12] Kshirsagar, V. K., Tidke, S. M., and Vishnu, S. Intrusion detection system using genetic algorithm and data mining: An overview. *International Journal of Computer Science and Informatics ISSN (PRINT) 2231*, 5292 (2012).

[13] Kudela, J. A critical problem in benchmarking and analysis of evolutionary computation methods. *Nature Machine Intelligence*, 4 (2022), 1238–1245.

[14] Li, L.-H., Ahmad, R., Tsai, W.-C., and Sharma, A. K. A feature selection based dnn

for intrusion detection system. In *2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM)* (2021), IEEE, pp. 1–8.

[15] Liu, Z., Liang, X., and Huang, M. Design of logistic regression health assessment model using novel quantum pso. In *2018 IEEE 3rd International Conference on Cloud Computing and Internet of Things (CCIOT)* (2018), IEEE, pp. 39–42.

[16] Maalouf, M. Logistic regression in data analysis: an overview. *International Journal of Data Analysis Techniques and Strategies 3*, 3 (2011), 281–299.

[17] Marso, S., and El Merouani, M. Predicting financial distress using hybrid feedforward neural network with cuckoo search algorithm. *Procedia Computer Science 170* (2020), 1134–1140.

[18] Muller, J. Improving initial aerofoil geometry using aerofoil particle swarm optimisation. In *Mendel* (2022), vol. 28, pp. 63–67.

[19] Ngatman, M. F., Sharif, J. M., and Ngadi, M. A. A study on modified pso algorithm in cloud computing. In *2017 6th ICT international student project conference (ICT-ISPC)* (2017), IEEE, pp. 1–4.

[20] Parmisano, A., Garcia, S., and Erquiaga, M. J. A labeled dataset with malicious and benign iot network traffic. *Stratosphere Laboratory: Praha, Czech Republic* (2020).

[21] Potluri, S., and Diedrich, C. Accelerated deep neural networks for enhanced intrusion detection system. In *2016 IEEE 21st international conference on emerging technologies and factory automation (ETFA)* (2016), IEEE, pp. 1–8.

[22] Qasim, O. S., and Algamal, Z. Y. Feature selection using particle swarm optimization-based logistic regression model. *Chemometrics and Intelligent Laboratory Systems 182* (2018), 41–46.

[23] Qi, B., Wu, M., and Zhang, L. A dnn-based object detection system on mobile cloud computing. In *2017 17th International Symposium on Communications and Information Technologies (ISCIT)* (2017), IEEE, pp. 1–6.

[24] Rachit, Bhatt, S., and Ragiri, P. R. Security trends in internet of things: A survey. *SN Applied Sciences 3* (2021), 1–14.

[25] Sahar, N., Mishra, R., and Kalam, S. Deep learning approach-based network intrusion detection system for fog-assisted iot. In *Proceedings of international conference on big data, machine learning and their applications: ICBMA 2019* (2021), Springer, pp. 39–50.

[26] Samizadeh Nikoui, T., Rahmani, A. M., Balador, A., and Haj Seyyed Javadi, H. Internet of things architecture challenges: A systematic review. *International Journal of Communication Systems 34*, 4 (2021), e4678.

[27] Sarker, I. H. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science 2*, 6 (2021), 420.

[28] Singh, G., and Khare, N. A survey of intrusion detection from the perspective of intrusion datasets and machine learning techniques. *International Journal of Computers and Applications 44*, 7 (2022), 659–669.

[29] Wang, K.-J., Makond, B., Chen, K.-H., and Wang, K.-M. A hybrid classifier combining smote with pso to estimate 5-year survivability of breast cancer patients. *Applied Soft Computing 20* (2014), 15–24.

[30] Wang, S., Balarezo, J. F., Kandeepan, S., Al-Hourani, A., Chavez, K. G., and Rubinstein, B. Machine learning in network anomaly detection: A survey. *IEEE Access 9* (2021), 152379–152396.

[31] Zhang, H., Shi, Y., Yang, X., and Zhou, R. A firefly algorithm modified support vector machine for the credit risk assessment of supply chain finance. *Research in International Business and Finance 58* (2021), 101482.

[32] Zhang, Y., Li, P., and Wang, X. Intrusion detection for iot based on improved genetic algorithm and deep belief network. *IEEE Access 7* (2019), 31711–31722.

[33] Zhang, Z. Speech feature selection and emotion recognition based on weighted binary cuckoo search. *Alexandria Engineering Journal 60*, 1 (2021), 1499–1507.