Int. J. Advanced Networking and Applications
Volume: 11 Issue: 06 Pages: 4475-4485 (2020) ISSN: 0975-0290

Optimal Feature Subset Selection Using Cuckoo Search on IoT Network

Samah Osama M. Kamel
Department of Informatics Research, Electronics Research Institute, Cairo, Egypt
Email: samah_n2003@yahoo.com
SanaaAbouElhamayed
Department of InformaticsResearch, ElectronicsResearch Institute, Cairo, Egypt
Email:sanaa-hamayed@hotmail.com

ABSTRACT

The Internet of Things (IoT) became the basic axis in the information and network technology to create a smart environment. To build such an environment; it needs to use some IoT simulators such as Cooja Simulator. Cooja simulator creates an IoT environment and produces an IoT routing dataset that contains normal and malicious motes. The IoT routing dataset may have redundant and noisy features. The feature selection can affect on the performance metrics of the learning model. The feature selection can reduce complexity and over-fitting problem. There are many approaches for feature selection especially meta-heuristic algorithms such as Cuckoo search (CS). This paper presented a proposed model for feature selection that is built on using a standard cuckoo search algorithm to select near-optimal or optimal features. A proposed model may modify the CS algorithm which has implemented using Dagging with base learner Bayesian Logistic Regression (BLR). It increases the speed of the CS algorithm and improves the performance of BLR. Support Vector Machine (SVM), Deep learning, and FURIA algorithms are used as classification techniques used to evaluate the performance metrics. The results have demonstrated that the algorithm proposed is more effective and competitive in terms of performance of classification and dimensionality reduction. It achieved high accuracy that is near to 98 % and low error that is about 1.5%.

Keywords – **IoT, Feature Selection, Meta-heuristic, Cuckoo Search Algorithm, Deep Learning.**

Date of Submission: May 12, 2020 Date of Acceptance: June 16, 2020

1. INTRODUCTION

The Internet of Things (IoT) plays an important role in the information technology revolution. IoT creates a new intelligent environment to build a smart world. IoTapplications improve people's life and all fields of communication, industries, education, commercial business, healthcare, agriculture, etc. The IoT network provides the connection between physical and virtual objects such as smart devices using sensors. Therefore; the IoT network takes advantage of every smart device and creates a new system with newly developed capabilities. All smart devices can send and receive all information among each other.

The structure of the IoT network contains six layers namely; coding, perception, network, middleware, application, and business layer. In this work; we focus on the network layer because it is the key element in the IoT network. The network layer is responsible for the specification and functions of physical objects. The network layer connects different smart devices to exchange data over the IEEE 802.15 protocol using IPv6. IPv6 protocol is responsible for the connection between smart devices via IPv6 over Low-Power Wireless Personal Area Networks (6LoWPAN) [2]. The routing process is based on Low Power and Lossy Networks (LLNs) named RPL which is used to build a network over constrained nodes [3]. Consequently; these devices and the exchanged data will be exposed to different types of attacks that give rise to the exhausting of network resources such as power and memory. Besides; these attacks can damage the entire network and have a bad effect on security requirements such as privacy, availability, integrity, and confidentiality.

The creation of a real environment of IoT network needs to test and correct new protocol especially RPL protocol and the properties of many smart sensors that is power consumption, performance, and memory. Therefore; there are many simulators to achieve such requirements for IoT networks. One such simulator is the Cooja Simulator that is the most famous simulator of the IoT network which creates a real environment of IoT network by generating many nodes including normal and malicious nodes. The identification of malicious nodes (attacks) needs to analyze strange behavior and their effect on power consumption. Thence; the detection (classification) technique with high-performance metrics is a major problem because the routing dataset has many irrelevant, noise, and redundant features. The high- dimensionality reduction of features affects onthe accuracy of the learning model when unwanted and insignificant features are used. Therefore; reducing these features is an important step to build a learning model with high-performance metrics. The dimensionality reduction can be divided into feature extraction and selection. The feature selection approaches are split into three parts, namely, flat feature, structure feature, and streaming feature. The flat feature technique is typically divided into three techniques that are filter, wrapper, and embedded. The filter feature technique depends on the characteristics of the training data such as distance, consistency, dependency, information, and correlation. It doesn't depend on that the selected feature will achieve high-performance metrics.

In contrast; the wrapper feature technique depends on the evaluation of the quality of the selected features which are used to increase performance metrics. Then the wrapper technique is more efficient than a filter feature technique to build robust learners to achieve high performance [4].

Most of the researches try to find solutions for hard optimization problems in the real-world which is implemented in a finite time. Therefore; meta-heuristic algorithms are used to solve the problem of heuristic optimization algorithm in a certain or small enough time. Meta-heuristic is a high-level problem which is an independent framework which provides a set of strategies schemas to develop optimization problem. The idea of meta-heuristic is nature-inspired that provides high performance to solve complex problems [5]. Meta-heuristic algorithms can solve a problem of large scale data with high dimensionality which is an obstacle to most researchers. It is a convenient solution for the global optimization problem [6]. Meta-heuristic is classified into two categories namely; biological algorithms and bio-inspired algorithms. Biological algorithms include simulated annealing, gravitational search algorithm, magnetic optimization algorithm, external optimization, and harmony search. Bio-inspired algorithms are divided into two classes that are evolutionary methods such as genetic algorithm, evolution strategy, evolutionary programming, and swarm intelligent algorithms. Swarm intelligent algorithms are inspired by organisms' life that is classified into Stigmergic-based algorithm and imitation-based algorithm. Stigmergic-based algorithm is used to interact among all elements of the environment indirect way such as Ant Colony Optimization. On the contrary; the imitation-based algorithm is used to interact among all elements of the environment indirectly way. The imitation-based techniques are particle swarm optimizer, firefly Algorithm, bat Algorithm, and Cuckoo search [7].

All these algorithms deal with high dimensionality data and can select the significant features and get rid of features which have no value in local and global search. Therefore; feature selection is a very important step because it makes the training process faster. Moreover; the feature selection process reduces the model complexity and builds a robust model with high-performance metrics. The feature selection process can reduce the over-fitting problem.

In this work, the wrapper feature technique is used for feature selection using a standard cuckoo search algorithm which is used for selecting an optimal or near-optimal feature subset. The standard cuckoo search (CS) algorithm has been implemented using Meta classifier Dagging with base learner Bayesian Logistic Regression (BLR). Dagging improves the performance rate of BLR. In addition to the implementation of BLR takes a long time.

The standard CS algorithm is better than other algorithms that it has the ability of local and global search so it is a more robust and efficient algorithm. Standard CS algorithm is based on L´evyflights which uses random walk to explore to search space to improve CS. The disadvantage of standard CS is that its convergence rate that it may be a little slower [17, 18].

Therefore; this paper introduced a proposed model that modifies the standard CS algorithm using Dagging based on BLR to obtain a robust model for the feature selection method. This proposed model improves the speed of the standard CS algorithm and the performance of BLR through utilizing a Dagging ensemble learner.This work achieves two experiments to present two models to introduce the impact of Dagging on standard CS algorithm and BLR. The first model has been implemented using standard CS algorithm using Dagging based on BLR and the second model has been performed using standard CS algorithm using BLR. Therefore; there are two models. The performance metics of the first model based on standard CS algorithm using Dagging based on BLR is higher than the model of standard CS algorithm using BLR. additionally; The using of Dagging based on BLR increases the speeed of the standard CS algorithm and improves the performance of BLR.

The remaining paper has been structured as follows: related works present some of the researches. Section three shows the system Architecture that illustrates IoT routing dataset generation, classification techniques, and the Proposed Optimized Approach. The performance metrics have been used in this work and can be displayed in section four. Section five looks at the methodology used to build the paper. The experimental results are introduced and analyzed in Section six. The final part is the conclusion for this work.

2. RELATED WORKS

The feature selection methods are the main axis for classification and prediction performance to obtain a robust model. In this section, many types of research presented recent stochastic methods for solving feature selection problems. There are many common contradictory criteria among all of these methods to improve the performance metrics through the enhancement of many algorithms to find the best solutions.

[18] The authors presented a new meta-heuristic optimization approach, called Parasitism-Predation Algorithm (PPA) for feature selection which was inspired by a combination of advantages of cat swarm optimization (CSO), cuckoo search (CS) and crow search algorithm (CSA) and solved different problems. A novel Parasitism-Predation Algorithm (PPA) simulated the interaction between the predator (cats), the parasite (cuckoos), and the host (crows). All these phases were suggested to develop improve exploration and exploitation capability. This work used the distribution and reinitialization strategies of L´evyflight to enhance exploration ability and augment the variety of the search space. PPA-based feature selection was used to decrease the number of features and select the most significant features. The base learner was KNN classifier. This study used benchmark real-world datasets to evaluate its performance and quality. The experimental

results had shown that the PPA optimization algorithm is suitable for a heuristic method for data classification problems to many problems. The results had indicated that PPA can avoid local minima compared to CS, CSA, CSA, and WOA algorithms for all tested datasets except ionosphere, glass, and multi-feature datasets. The performance of the selected features achieved high accuracy but PPA had an obstacle that it consumed time.

[20] This study was based on the application of BCS on the Kaggle dataset for optimized feature selection supervised soft computing techniques. The authors used feature selection optimization using the binary Cuckoo search algorithm. The implementation of this study was based on anumber of tweets (7086 tweets) that wereanalyzed based on accuracy. The idea of this study was based on a description that only more than 53.17% of the feature selected can achieve the average accuracy gain is of 7.45%. This work used tf-idf technique for feature selection and the result is that 47% of features were noice and redundant that affected accuracy. This method increased over-fitting problems. the results had shown that Support Vector Machine (SVM), Multilayer Perceptron (MLP), NaïveBayesian (NB), DecisionTree (DT) achieved high accuracy but K-Nearest Neighbour (KNN) achieved the lowest accuracy.

[24] This work proposed a Reinforced Cuckoo Search Algorithm (RCSA) for multimodal optimization, which consisted of three different strategies: modified selection strategy, Patron-Prophet concept, and self-adaptive strategy. The modified selection strategy was based on a standard Cuckoo Search (CS) to obtain an efficient selection of next-generation individuals in the state of selecting a random set of individuals. The Patron-Prophet concept had been defined to extract the deviated information from the used individual. It used to generate new individuals to perform exploitation. The last strategy is a self-adaptive step size parameter 'α' that had been renamed from static to dynamic for maintaining the diversity during the search. The proposed algorithm had been estimated and compared with the LSGO with benchmark instances with 500 and 1000 dimensions and three engineering design problems. The proposed model had been estimated using minimum, mean, and standard deviation, the Friedman Average Ranking test (statistical analysis), convergence analysis, runtimeanalysis; categorical behavior the proposed RCSA achieved high-performance metrics.

[21] The authors introduced a proposed modified cuckoo search algorithm with a rough set to reduce the number of features and achieve high accuracy. The Experiments had been implemented over many different datasets. The obtained results had shown that the proposed algorithm significantly had been performed better than other algorithms to reduce irrelevant, redundant, or noisy features and achieved high accuracy. The proposed model was more stable than other methods and can generate feature subsets.

[17] The authors presented a proposed Extended Binary Cuckoo Search optimization algorithm aims to improve the feature selection. The target of the proposed Extended Binary Cuckoo Search algorithm was to minimize the number of selected features or to select the most significant features to increase accuracy. The proposed model presented a new objective function which deems the number of selected features as the regulation term and classification accuracy to determine the best subset of features. In the proposed model; each subset of feature is input into the SVM classifier. Experiments were implemented to compare the Extended Binary Cuckoo Search against other meta-heuristic algorithms, whichwere the Binary Ant Colony Optimisation, Binary Genetic Algorithm, and Binary Particle Swarm Optimisation. The proposed model was based on using biomedical datasets. The results had shown that the proposed algorithm was better than all other algorithms for selecting the most significant features. The proposed model decreased the number of features and achieved higher accuracy and higher speed than other algorithms. Therefore; it could reduce the complexity, time, and computational processing power.

[19] The authors developed a new algorithm hybrid Binary Cuckoo Search and rough set theory (FS-BCS) for classification on nominal datasets to select the best features to obtain high accuracy. The benchmark dataset was fice nominal datasets from the UCI repository which were Breast-W, Mushroom, Dermatology, Soybean (large), and Chess. The experimental results had shown that the proposed algorithm had achieved better feature selection methods than NIAs such as ACO, ABC, and PSO. The proposed algorithm decreased the number of iterations and high accuracy. It had been implemented quickly and was more efficient convergence than other NIAs. FS-BCS decreased complexity and few parameters. FS-BCS achieved higher SR% and improved the DT and NB classification accuracy compared to the Genetic and PSO approaches. The experimental results had demonstrated that PSO hadn't succeeded to do so in two datasets and Genetic failed in three datasets. The results had shown that BCS with RSTDD has faster and more efficient convergence compared to PSO and Genetic approaches

[27] This research accounted for a multi-objective cuckoo search based on using a correlation-based filter model to feature selection. The proposed system was estimated using different measures over the different data sets. The results were compared against the results of genetic and particle swarm optimization with single and multiple objectives. The experimental results had shown that the proposed model can decrease the number of features and the selected features with minimum correlation and achieved high performance. The proposed multi-objective system was more robust and stable than other algorithms.

[26] This work addressed the feature selection of cancer classification using gene expression data. The proposed model had achieved two levels; first using the Cuckoo search algorithm to search the information genes from the top-m ranked genes and for classification. The proposed model was based on the fitness function for CS. k-Nearest Neighbour (kNN) technique was used to evaluate the classification accuracy. The results had illustrated that the

CS-based method achieves 100% accuracy. The benchmark datasets were DLBCL Harvard, Lung Michigan, Ovarian, AML-ALL, and Lung Harvard2 datasets.

[38] The authors proposed a new New Filter Feature Selection Method based on Binary Cuckoo Optimization Algorithm (BCOA). This proposed algorithm was based on BCOA and the Mutual Information (Mi) (entropy and mutual information) of each pair of features to specific relevance and redundancy of the selected feature Subset. The main target was to select the most important features for high dimensional datasets and achieve high-performance metrics. The experiments had been evaluated by An Artificial Neural Network (Ann) using six datasets. The experimental results had demonstrated that the proposed model can reduce the number of features and accomplish high accuracy.

[25] This work introduced a binary version of the Cuckoo Search (BCS) with various transfer functions that map continuous solutions to binary ones to select an important feature to achieve high-performance metrics. This paper had set the parameters dynamically to improve the performance of Binary Cuckoo Search to get robustness of the CS algorithm. The experiments had been implemented and compared BCS with the binary version of the Bat algorithm. The proposed model was based on four public datasets and had been applied using cross-validation. The experimental results had shown that the binary cuckoo version keeps the parameters α and pa fixed during all iterations. Therefore; BCS had implemented efficiently to reduce the number of features and achieved high accuracy.

[23] This paper introduced an adaptive Cuckoo search algorithm for optimization to fulfill dynamically increasing switching parameters to improve CS algorithm performance. This proposed model was based on using linear, exponential, and power of the CS algorithm to increase switching parameters and tested against the constant and linearly decreasing CS algorithms. This experiment increased the exponential increasing (CSEI) switching parameter to be more efficient than other CS algorithms.

[28] A evolutionary mechanism improved the CSDE which combined the L´evy fight of CS with a modified mutation operation of DE dynamically to generate a suitable solution. The mutation operation of DE changed under different searching stages. This proposed model used 10 benchmarking to solve a logistics distribution center. The experimental results had shown that the performance of the CSDE is the best solution, mean solution, and convergence speed than several meta-heuristic algorithms. In addition to, the CSDE had performed effectively better than or equal to CS, ICS, and some other meta-heuristic algorithms.

3. THE SYSTEM ARCHITECTURE

Themain step is creating an IoT environment using the Cooja simulator to generate various nodes involve normal and malicious nodes. The selection and position of these nodes are random. The next process is feature extraction to select the most important feature to increase the performance metrics. The final process is the evaluation of the proposed model to build a robust classifier model using different classification algorithms.

3.1. IoT Routing Dataset Generation

The CoojaIoT simulator for Contiki3.0 OS operating system is used to implement the simulation of different IoT network that is based on the RPL protocol. Cooja simulator is run C programming language. The experiment has implemented using a virtual machine (workstation 15) with 16 GB RAM and an operating system 64-bit Ubuntu (Ubuntu 16). http://www.contiki-os.org/

The second step is gathering an IoT routing dataset that consists of 5577 motes (nodes) including normal motes (1) and malicious motes (0). The dataset seems a small dataset so Synthetic Minority Over-Sampling Technique (SMOTE) is used to increase the size of the IoT routing dataset and becomes 17092 motes [39]. The IoT routing dataset is split in the ration of 80:20 (training-testing) with 10-fold cross-validation. The training datacontains13674 and the test 3418 including normal motes (1) and malicious motes (0). The IoT routing dataset contains 20 features and one label that indicate normal (1) and malicious (0). The generated features are Node, Received, Dups, Lost, Hops, Rtmetric, ETX, Churn, Beacon Interval, CPU Power, LPM Power, Listen to Power, Transmit Power, Power, On-Time, Listento Duty Cycle, Transmit Duty Cycle, Avg. Inter Packet, Min Inter Packet, Max Inter Packet, and Label.

3.2. Classification Technique

SupportVector Machine (SVM) is based on the idea that a hyperplane in feature space a decision surface is constructed to map input vector (classes) to a high dimensional feature space to differentiate the classes. The advantage of the decision surface is that ensuresa high generalization capability of the learning machine. In hyperspace, the labeled training set is designed as points. -1 hyperplane will be calculated when n classes of the training dataset are classified [29].

As we mentioned above, the IoT environment generates different types of data because of the nature of the IoT network where various types of smart devices are used to build the IoT network. Therefore; SVMs are suitable algorithms to deal with different types of data that have unstructured data. Besides; SVMs are used for high dimensional data and decrease over-fitting.

The hyperplane can be written as follows:

$$w^{\rightarrow}.x^{\rightarrow} - b = 0 \qquad (1)$$

Where:w is the vector, x is data input and b is the constant that is labeled from the training dataset.

There are many types of SVM but the proposed model is used linear SVM because of its speed and ability to learn new datasets. As we mentioned above; the nature of the IoT environment is a dynamic system and IoT receives many data from different smart devices so the linear SVM is used.

Deep learning is a type of artificial intelligence approach which consists of an input layer, multiple hidden layers, and the output layer. Deep learning gains much popularity to get high accuracy when a huge amount of data is used. Deep learning may be flexible and very domain-specific. The convolution neural network (CNN) is a common deep learning algorithm and is suitable for any type of data. CNN consists of a convolution layer, batch normalization layer, activation layer, pooling layer, regularization layer, and fully connected layer [30]. The major advantage of CNN is that it automatically detects the most significant features without any human interaction. The convolution and pooling operations contribute making CNN computationally efficient. The process of CNN can be described as the following steps: -

Algorithm Learning Model based on CNN

Input
IoT routing dataset
Process
For I =1 to N do
For each batch of Bm, input data do
Compute the results of the convolutional layer$f(x_i, w_i, b) = \sum_i w_i x_i + b$
Compute the activation function (ReLU)The activation function size $= 1 + \frac{N-F}{S}$
Implement the max-pooling process$f(x) = \max(0, x)$
Compute result of the fully connected layer with a softmax function$S(y_i) = \frac{\exp^{y^j}}{\sum_j^i \exp^{y^i}}$
Compute errors by using the loss function based on cross-entropy between the output value and the label value $H(p, q) = \sum_x p(x) \log q(x)$
End for
End

Where: x_i is an instance, w is the weight of instance and b is bias, N is window size, F is filter size and S is stride, y^j is the output of the j^{th} neuron.
Fuzzy Unordered Rule Induction Algorithm or FURIA algorithm is modified and extended to the RIPPER classifier [31]. FURIA deals with a problem of fuzzy rule-based classification system that combinesthe merits of the RIPPER algorithm with fuzzy logic. RURIA algorithm starts with fuzzification of rule predecessors that is provided by RIPPER which passing from crisp intervals to trapezoidal fuzzy sets. Therefore; the authentic crisp intervals define the core of the new fuzzy sets however supports are expanded to try to get the maximum of the data instances coverage that is compatible with rule output. So predecessors are arranged according to their prorated significance [32]. FURIA is a simple and intelligible rule sets. FURIA algorithm deals with exposed data by using an efficient rule. The experimental results illustrate that the classification accuracy of the FURIA algorithm is significantly high in the original RIPPER. The following equation is the final format of FURIA.

$$Ri: \quad IF\ X_1\ is\ A_i^i\ AND\ X_n\ is\ A_n^i$$
$$Then\ Y\ is\ B^i(w^i)$$

Where: Ri is the rule, B^i is the rule output class, W^i is the weight associated to rule, A^h is the rule of predecessors for variable X^h that $(h \in [1, n])$, $A_h^i = \{a_h^1, a_h^2, a_h^3, a_h^4\}$, presents the four points which are characteristics of a trapezoidal fuzzy set.
The parameters a_h^2, a_h^3 are bound of the core $(\mu A_h^i(x) = 1\ if\ x \epsilon\ [a_h^2, a_h^3]$ whilea_h^1 , a_h^4 limiting the support $A_h^i(x) > 0\ if\ x \epsilon\ [a_h^1, a_h^4]$.

The rules are weighted according to their Certainty Factor (CF) which is used to reflect the proportion of data instances that are covered correctly. Once the instance is covered by the set of produced rules, FURIA predicts the class that is induced from the sum of activation degree of all produced rules per class. Else, a new set of rules is created dynamically by FURIA that takes the merit of rule stretching mechanism. Else, a new set of rules is created dynamically by FURIA that takes the merit of rule stretching mechanism. FURIA checks rule by rule all induced rules set for a given instance. For each produced rule, predecessors are deleted from the least to the most important one. When the instance is covered by the stretched rule it passes to the analysis of the next rule [32].

3.3. The Proposed Optimized Approach

In this work, the wrapper feature technique is based on using a standard cuckoo search algorithm that has been implemented through Dagging with base learner BLR to enhancement the routing path among motes to differentiate between normal routing path and malicious path. This method is implemented to select the most significant features and get rid of unimportant features. The experiment uses the 10-fold cross-validation approach to divide data into training and testing sets.
The standard CS algorithm is the most modernistic meta-heuristic s algorithm. The idea of a standard CS algorithm is based on the parasite behavior of cuckoo birds and their nets. The main parameters in this algorithm are birds and nets and the mechanism of motion or walk. Therefore; the philosophy of the standard CS algorithm is built on cuckoo breeding behavior and Lévy flight. Regards cuckoo breeding behavior, the standard CS algorithm presumes that each bird can produce eggs and try to search to lay eggs on hosts' nets. Each bird selects a host or nest to a position in the nest randomly. The number of the nest is fixed and can't be changed over time. In case the nests have the best quality eggs they will be the best nets. The best nets will be the best selection for the nest newborn bird. If a host bird will detect the bird left behind by cuckoo it can either give up the nest or build a new nest. So the probability of those eggs of the cuckoo will detect the host bird can be expressed asPa ∈ [0,1]. This algorithm combines the local random walk and global random walk which are controlled by Pa [25]. The local random walk can be expressed as follows:

$x_i^j(t) = x_i^j(t-1) + \propto \cdot s \oplus H(Pa - \mathcal{E}) \oplus (x_{k'}^j(t-1) - x_{k''}^j(t-1))$ (2)

Where: x_i^j number of eggs j in nest i that j=1,2,……,d and i=1,2,…..,m.

$x_{k'}^j$ and $x_{k''}^j$ are two different solutions that are selected by random alteration.

H(·) is a Heaviside function, $\mathcal{E}$ is random n drawn from a uniform distribution, and s is the step size.

The global random walk is performed using Lévy flights that can be noted as follows:

$$x_i^j(t) = x_i^j(t-1) + \propto \cdot L(s,\lambda) \quad (3)$$

$$L(s,\lambda) = \frac{\lambda \cdot \Gamma(\lambda) \cdot \sin(\lambda)}{\pi} \cdot \frac{1}{s^{1+\lambda}}, s \gg s_0 > 0 \quad (4)$$

Lévy flight provides a random walk Lévy distribution set a random step length as illustrated:

$$\frac{1}{s^{\lambda+1}} \quad (0 < \lambda \leq 2) \quad (5)$$

L´evyRandom walk implements many steps that have lengths and are distributed according to a heavy tail. These all steps help to improve a random walk and generate many solutions to produce the best or optimal solution or near-optimal solution to increase the local search speed [18].

Dagging is used to build Meta classifiers to improve the classification or prediction accuracy of the base learner (classifier). The task of Dagging is to generate and derived M dataset from the original training dataset. M dataset contains a number of samples N. The samples in each dataset differentiate from others where there is no two datasets have the same samples. Each dataset is used to train a base learner. After the raining process, the classifiers model is created. The final decision is obtained from using majority voting of the classification models [15, 34]. The Base learner of Dagging is BLR which causes an over-fitting dataset so the performance metrics decrease.

The logistic regression can't handle a high dimensional dataset. Therefore; the Bayesian approach is applied to logistic regression to overcome the problem of logistic regression. The idea of BLR is to use a prior probability distribution over predictor variables. It takes the nonlinear relationship between the predictor variables and the result variable using the Bayesian approach [35]. The sample of dataset posterior probability can be calculated as follows:

$$P(c|a_1, a_2, \ldots\ldots\ldots. a_n) = \frac{1}{(1 + e^{(b + w_0 * c + \sum_{i=1}^{n} w_i * f(a_i))})} \quad (6)$$

$$c = \log\frac{P(class=0)}{P(class=1)} \quad (7)$$

$$f(a_i) = \log(\frac{P(a_i|class=0)}{P(a_i|class=1)} \quad (8)$$

where: $a_1$is the predictor variables, c is the prior log odds ratio, b is the bais, $w_o$and$w_i$are weights that are learned from the training dataset and $f(a_i)$are the feature of i^{th} feature a_i.

Two parameters are important in the BLR namely; a univariate Gaussian prior with a mean '0' and variance of 'σi' over the weights. If a mean '0' is used the weights are near to zero. If the positive values of σi with small values it indicates that the values of weights are trusted. If the large values are point out the values are not important [35]. The most flaw of BLR is the over-fitting dataset so the performance is decreased. To enhance the performance of Bayesian Logistic Regression, the Dagging is used to improve the performance of BLR by reducing the over-fitting problem.

Dagging is applied to BLR to improve the performance metrics by decreasing the number of iteration and avoid over-fitting. So the speed of BLR will increase and accuracy decreases. The usage of Dagging with BLR as a base learner increases the speed of the CS algorithm as a search strategy and produces near-optimal or optimal features. Thence; Dagging Meta classifier is used to create model verification of BLR and increase the search methodology of the CS algorithm.

By using Dagging, it can decrease the number iteration of BLR and improve performance of BLR to avoid the over-fitting problem.

4. PERFORMANCE METRICS

The performance of the learning model can be evaluated by considering a set of measurable criteria. The performance metrics are used to evaluate the strength of the model. The criteria or the performance metrics are accuracy, error, and F-measure.

Accuracy: accuracy is the ratio of the number of correct predictions and the total number prediction. Formally, the equation of accuracy can be represented as follows

$$\mathrm{AC} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \quad (9)$$

Where: TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

Error: the classification error can be defined as the ratio between incorrect predictions and the total number prediction as follows.

$$\mathrm{Error} = \frac{\mathrm{FP+FN}}{\mathrm{TP+TN+FP+FN}} \quad (10)$$

F-measure or F1 score: it transfers the balance between the precision and the recall.

$$\mathrm{F-measure} = 2 * \frac{(\mathrm{Precision} * \mathrm{recall})}{(\mathrm{Precision} + \mathrm{recall})} \quad (11)$$

5. METHODOLOGY OF THE PROPOSED MODEL

As mentioned before, the prediction of IoT routing attacks is one of the significant challenges of IoT networks. To improve classification performance, the most important features should be selected and get rid of irrelevant and noise features.

The experiment is based on the wrapper model that it contains three stages which are generating a subset, learning algorithm, and performance process. This section presents a proposed method based on selecting the features. The idea of the proposed method can be illustrated as shown in the flowing figure.

The flow chart of the proposed method starts with generating of IoT routing dataset using Cooja Simulator. The next step is the pre-processing IoT routing dataset using the SMOTE technique to increase the size of the generating dataset. The next process is dividing the IoT routing dataset vertically into four blocks. Each block contains the IoT routing dataset with specific features.

Each block contains five features that are selected randomly. The CS algorithm is applied to each block and producesthe best features that give the best performance metrics.

Dagging with base learner BLR is used for the classification process to evaluate the selected features from each block. The following step is generating model validation. If the termination criteria are satisfied the best feature is obtained. Otherwise, the process repeats many times randomly to get the most significant features and achieve high accuracy. It is notable that some features have a high correlation with each other's but they can't achieve high performance. Therefore; we select the features which produce high performance and have good correlations.

There are two experiments to create two models. The first model has performed using CS algorithm using Dagging with base learner BLR and the second model has implemented using CS algorithm using BLR.

The purpose of two experiments is to show the impact of Dagging in the feature selection and the performance metrics to build the training model using classification techniques. Dagging enhances the performance of BLR by decreasing the number of iterations. The number of iterations of BLR decreases when we use Dagging. The speed of standard CS algorithm is increased with minimum number of iterations.

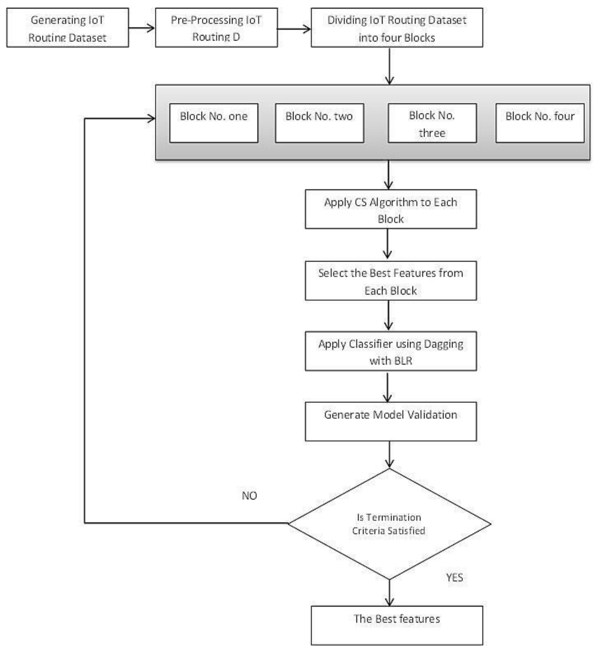

Fig.1.The flowchart of the proposed Method.

6. EXPERIMENTAL RESULTS AND DISCUSSION

As we mentioned above, the CoojaIoT simulator for Contiki3.0 OS operating system is used to build an IoT environment and generates IoT routing datasets. The experiments have been implemented using the platform that is Intel Core i7 4500CPU, 2.40GHz, and RAM is 8GB and MS windows 10 professional 64 bits. The Waikato Environment for Knowledge Analysis (Weka) version 3.7.12 is used to implement the experiments which are an open-source machine learning package. Weka simulator contains many processes that are the Explorer, Experimenter, Knowledge Flow, and Simple CLI. For the Explorer section, it contains tools for data pre-processing, classification, clustering, association rules, select attributes, and visualization.

This work presents the effect of the CS algorithm using Dagging with BLR as a base learner to select the most significant features to enhance the performance metrics. As we mentioned before; the Cooja simulator is used to generate an IoT routing dataset that contains normal malicious nodes. Some features describe the IoT routing dataset. It is noted that; there are some feature isn't important and cause noise and over-fitting problem and they have a bad effect on the performance metrics. The CS algorithm is a modern approach that helps to select the near-optimal or optimal features. Using Dagging with base learner BLR has a contribution to select the most important feature to increase the performance metrics.

After using the feature selection process, the classification algorithms are used to evaluate the performance metrics to illustrate how the selected features have an impact on the performance metrics.

The IoT routing dataset is divided into four blocks randomly. We have four blocks of features so the first experiment has been performed using CS algorithm using Dagging with base learner BLR. Table 1 presents the selected features of each block.

Table 1. Four Blocks of the Selected Features using CS Algorithm using Daging with BLR.

Blocks	Features
First Block	3 (Lost), 5 (Rtmetric), 6 (ETX)
Second Block	7 (Churnn), 10 (LPM Power), 12 (Transmit Power)
Third Block	13 (Power), 14 (On-time), 15 (Listen Duty Cycle)
Fourth Block	16 (Transmit Duty cycle), 17 (Avg. Inter-packet , 19 (Max Inter-packet)

From Table 1; the first model produced the selected features that are Lost, Rtmetric, ETX, Churn, LPM Power, Transmit Power, Power, On-time, Listen Duty Cycle, Transmit Duty cycle, Avg. Inter-packet, Max Inter-packet, and Label.

The performance metrics of the classification techniques after using the feature selection using the CS algorithm using Dagging with base learner BLR are represented in the following table.

Table 2. The Performance Metrics of the Classification Techniques after using CS Algorithm using Dagging with base learner BLR.

Classification Algorithms	Accuracy	Error	F-measure
SVM	96.02	3.98	91.49
CNN	98.57	1.43	98.32
FURIA	98.09	1.91	97.15

The second experiment is used to produce the second model which has been performed using the CS algorithm using BLR and produced the selected features. The selected features are Received, Dups, Lost, Rtmetric, ETX, Churn, Beacon, PM Power,Listen Power, Transmit Power, Listen Duty Cycle, Transmit Duty cycle, Avg. Inter-packet, Mini Inter- packet, Max Inter-packet. Table 2 introduces the selected features of each block.

Table. 3. Four Blocks of the Selected Features using CS Algorithm using BLR.

Blocks	Features
First Block	1 (Received), 2 (Dups), 3 (Lost), 5 (Rtmetric)
Second Block	6 (ETX), 7 (Churn), 8 (Beacon), 10 (LPM Power),
Third Block	11 (Listen Power), 12 (Transmit Power), 15 (Listen Duty Cycle), 16 (Transmit Duty cycle)
Fourth Block	17 (Avg. Inter-packet), 18 (Mini Inter-packet) , 19 (Max Inter-packet)

The performance metrics of the classification techniques after using the feature selection using the CS algorithm using BLR are represented in the following table.

Table. 4. The Performance Metrics of the Classification Techniques after using the CS Algorithm using BLR.

Classification Algorithms	Accuracy	Error	F-measure
SVM	91.02	8.98	84.2
NN	92.57	7.43	90.4
URIA	91.89	8.11	89.7

The results illustrate that the accuracy of using the CS algorithm using Dagging with base learner BLR (first model) is better than the accuracy of using the CS algorithm using BLR (second model).
The first experiment has been implemented using decreasing the number of iterations of the CS and using Dagging with base learner BLR. Where the number of iterations of BLR is decreased and Dagging compensates the number of iterations reduction of BLR.
The contribution of the first experiment is that the speed of the CS algorithm is increased by decreasing the number of iterations of the CS algorithm by using Dagging. The objective of using Dagging increases the speed of the CS algorithm and decreased the number of iterations of BLR. At the same time; the performance of BLR is increased.

The first model achieved high accuracy, low error, and high F-measure. The accuracy of the second model is lower than the accuracy of the first model because of the number of iterations. The speed of the CS algorithm in the second model is lower than the first model. It is obvious that the difference between using Dagging in the first experiment and the second experiment.
The following figure shows the accuracy rate of the classification techniques of two experiments (first and second models).
The first model achieved the accuracy of the classification techniques is higher than the accuracy of the classification techniques in the second modelas shown in the following figure.
Where: the accuracy of the first model using SVM, CNN and FURIA is SVM1, CNN1 and FURIA1. The accuracy of the second model using SVM, CNN and FURIA is SVM2, CNN2 and FURIA2.

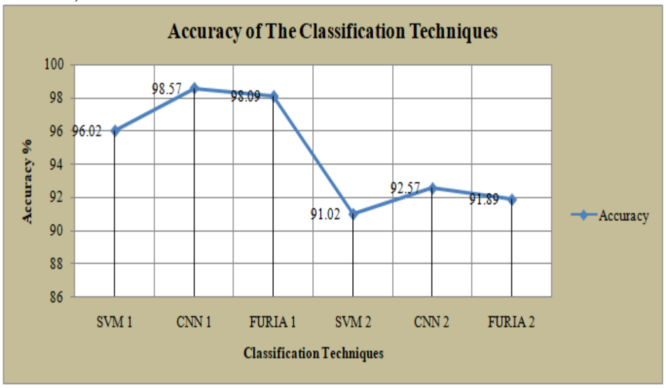

Fig. 2. The accuracy of the Classification Techniques of Two Experiments

The error of the classification techniques that used in the first model is higher than the error of the classification techniques used in the secondmodel as shown in the following figure.
Where: the error of the first model using SVM, CNN and FURIA is SVM1, CNN1 and FURIA1. The error of the second model using SVM, CNN and FURIA is SVM2, CNN2 and FURIA2.

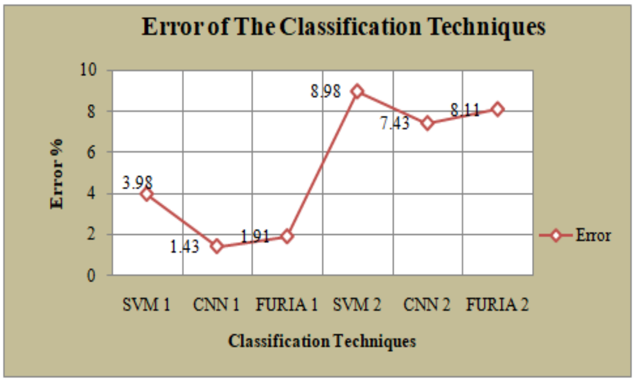

Fig. 3. The Error of the Classification Techniques of Two Experiments

The F-measure of the classification techniques used in the first model is higher than the F-measure of the classification techniques used in the second modelas shown in the following figure.
Where: the F-measure of the first model using SVM, CNN and FURIA is SVM1, CNN1 and FURIA1. The F-measure of the second model using SVM, CNN and FURIA is SVM2, CNN2 and FURIA2.

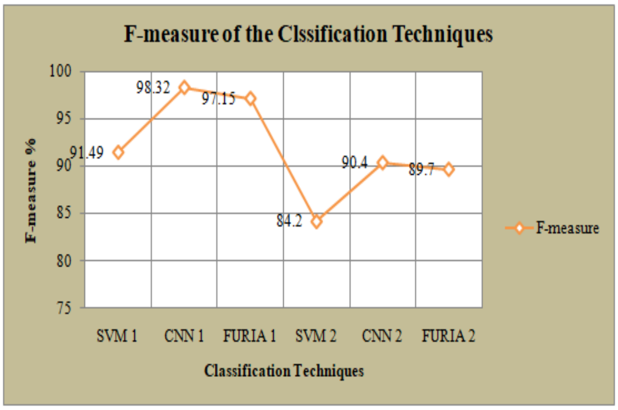

Fig. 4. The F-measure of the Classification Techniques of Two Experiments

The following figureshows the differences of performance metrics of the two models. The first model achieved higher performance metrics than the second model.

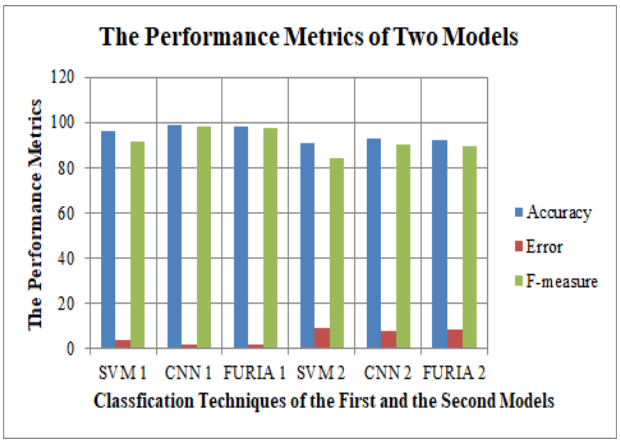

Fig. 5.The Performance Metrics of the Two Models.

7. CONCLUSION

This work analyzed and investigated the feature selection of the IoT routing dataset using a proposed model that was built on the CS algorithm using Dagging with base learner BLR. The methods were operated and tested using a dataset of IoT routing dataset that contained normal and malicious motes. There were two experiments to create two modelthat the first model had been implemented using CS algorithm based on Dagging with base learner BLR. The second model had been performed using the CS algorithm based on BLR. The results had shown that the Dagging ensemble learner increases the speed of the CS algorithm and improve the performance of BLR. The results had demonstrated that the proposed model is effective and competitive in terms of performance of classification and dimensionality reduction. It achieved high accuracy that is near to 98 % and low error that is about 1.5%. The accuracy of the proposed method outperforms the other adopted approaches. The proposed introduced better performance after rejecting any outlier instances than the dataset without cleaning. The accuracy value had changed by changing the number of selected features. The proposed model was more effective and robust in terms of performance of classification and dimensionality reduction.

REFERENCES

[1] G. Al-Rawashdeh, R. Mamat, and N. HafhizahBintiAbd Rahim, Hybrid Water Cycle Optimization Algorithm with Simulated Annealing for Spam E-mail Detection, *IEEE Access*, *7*, 2019, 143721 – 143734. DOI: 10.1109/ACCESS.2019.2944089

[2] A. Verma and V. Ranga, Analysis of Routing Attacks on RPL based 6LoWPAN Networks, *International Journal of Grid and Distributed Computing,* 2018, *11*(8), 2018, 43-56. DOI: 10.14257/ijgdc.2018.11.8.05.

[3] A. Mayzaud, R. Badonnel, and I. Chrisment, A taxonomy of attacks in RPL-based internet of things, *International Journal of Network Security*, *18*(3), 2016, 459-473.

[4] J. C. Ang, A. Mirzal, H. Haron, and H. N. A. Hamed, Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection, *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 13*(5), 2016. DOI: 10.1109/TCBB.2015.2478454.

[5] I. Boussaïd, J. Lepagnot, and P. Siarry, A survey on optimization meta-heuristic s, *Information Sciences, 237* 2013, 82–117. DOI: 10.1016/j.ins.2013.02.041.

[6] I. Kahvazadeh and M. SanieeAbadeh, MOCANAR: A Multi-Objective Cuckoo Search Algorithm for Numeric Association Rule Discovery, 2015. DOI: 10.5121/csit.2015.51509.99–113

[7] K. G. Dhal, S. Ray, A. Das, and S. Das, A Survey on Nature-Inspired Optimization Algorithms and Their Application in Image Enhancement Domain, *Archives of Computational Methods in Engineering,* 2019. DOI: 10.1007/s11831-018-9289-9.

[8] X. S. Yang, Nature-Inspired Optimization Algorithms, (Elsevier, 2014).

[9] Hira, Z. M. and Gillies, D. F, A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data, *Advances in Bioinformatics*, 2015, 1–13. DOI:10.1155/2015/198363

[10] C. Blum, J. Puchinger, G. R. Raidl, and A. Roli, Hybrid meta-heuristic s in combinatorial optimization: A survey, *Applied Soft Computing Journal,* *11*, 2011, 4135-4151. DOI: 10.1016/j.asoc.2011.02.032.

[11] A. Soler-Dominguez, A. A. Juan, and R. Kizys, A survey on financial applications of meta-heuristic s, *ACM Computing Surveys, 50*(1). 2017. DOI:10.1145/3054133.

[12] D. Jain and V. Singh, Feature selection and classification systems for chronic disease prediction: A review, *Egyptian Informatics Journal, 19,* 2018, 179-189. DOI: 10.1016/j.eij.2018.03.002.

[13] A. Safiyari and R. Javidan, Predicting lung cancer survivability using ensemble learning methods, *Ptoc. 2017 Intelligent Systems Conference, IntelliSys 2017, London, UK,* 7-8 Sept. 2017. DOI: 10.1109/IntelliSys.2017.8324368.

[14] N. Meddouri, H. Khoufi, and M. Maddouri, Parallel learning and classification for rules based on formal concepts, *Procedia Computer Science, 35,* 2014, 358-367. DOI: 10.1016/j.procs.2014.08.116.

[15] S. Wang, D. Wang, J. Li, T. Huang, and Y. D. Cai, Identification and analysis of the cleavage site in a signal peptide using SMOTE, dagging, and feature selection methods*, Molecular Omics*, *14*(1), 2018, 64–73. DOI: 10.1039/c7mo00030h.

[16] D. S. Anyfantis, M. G. Karagiannopoulos, S. B. Kotsiantis, and P. E. Pintelas, Local dagging of decision stumps for regression and classification problems, *Proc. 2007 Mediterranean Conference on Control and Automation, MED*, 2007. DOI: 10.1109/MED.2007.4433917.

[17] S. Salesi and G. Cosma, A novel extended binary cuckoo search algorithm for feature selection, *Proc. 2017 2nd International Conference on Knowledge Engineering and Applications, ICKEA 2017*, 2017. DOI:10.1109/ICKEA.2017.8169893.

[18] A. A. A. Mohamed, S. A. Hassan, A. M. Hemeida, S. Alkhalaf, M. M. M. Mahmoud, and A. M. Baha Eldin, Parasitism – Predation algorithm (PPA): A novel approach for feature selection, *Ain Shams Engineering Journal,* 2019. DOI:10.1016/j.asej.2019.10.004.

[19] A. F. Alia and A. Taweel, Feature Selection based on Hybrid Binary Cuckoo Search and Rough Set Theory in Classification for Nominal Datasets, *International Journal of Information Technology and Computer Science, 9,* 2017. DOI:10.5815/ijitcs.2017.04.08.

[20] A. Kumar, A. Jaiswal, S. Garg, S. Verma, and S. Kumar, Sentiment Analysis Using Cuckoo Search for Optimized Feature Selection on Kaggle Tweets, *International Journal of Information Retrieval Research, 9,* 2018. DOI:10.4018/ijirr.2019010101.

[21] M. A. El Aziz and A. E. Hassanien, Modified cuckoo search algorithm with rough sets for feature selection, *Neural Computing and Applications, 29,* 2018, 925–934. DOI:10.1007/s00521-016-2473-7.

[22] R. . Babukartik, Hybrid Algorithm using the advantage of ACO and Cuckoo Search for Job Scheduling, *International Journal of Information Technology Convergence and Services, 2,* 2012. DOI: 10.5121/ijitcs.2012.2403.

[23] M. Mareli and B. Twala, An adaptive Cuckoo search algorithm for optimisation, *Applied Computing and Informatics, 14*, 2018, 107-115. DOI: 10.1016/j.aci.2017.09.001.

[24] K. Thirugnanasambandam, S. Prakash, V. Subramanian, S. Pothula, and V. Thirumal, Reinforced cuckoo search algorithm-based multimodal optimization*, Applied Intelligence, 49,* 2019, 2059–2083. DOI: 10.1007/s10489-018-1355-3.

[25] L. A. M. Pereira et al., A binary cuckoo search and its application for feature selection, *Studies in Computational Intelligence,* 2014, 141-154. DOI: 10.1007/978-3-319-02141-6_7.

[26] C. Gunavathi and K. Premalatha, Cuckoo search optimisation for feature selection in cancer classification: A new approach, *International Journal of Data Mining and Bioinformatics, 13*(3), 2015, 248-65. DOI: 10.1504/IJDMB.2015.072092.

[27] W. Yamany, N. El-Bendary, A. E. Hassanien, and E. Emary, Multi-Objective Cuckoo Search Optimization for Dimensionality Reduction, *Procedia Computer Science, 96*, 2016, 207-215. DOI: 10.1016/j.procs.2016.08.130.

[28] R. Chi, Y. Su, Z. Qu, and X. Chi, A Hybridization of Cuckoo Search and Differential Evolution for the Logistics Distribution Center Location Problem, *Mathematical Problems in Engineering, 2019*, 2019. DOI: 10.1155/2019/7051248.

[29] M. V. Kotpalliwar and R. Wajgi, Classification of attacks using support vector machine (SVM) on KDDCUP'99 IDS database, Proc. 2015 5th International Conference on Communication Systems and Network Technologies, CSNT 2015, 2015. DOI: 10.1109/CSNT.2015.185.

[30] S. Albawi, T. A. Mohammed, and S. Al-Zawi, Understanding of a convolutional neural network, *Proc. 2017 International Conference on Engineering and Technology, ICET 2017,* 2018. DOI: 10.1109/ICEngTechnol.2017.8308186.

[31] A. Palacios, L. Sánchez, I. Couso, and S. Destercke, An extension of the FURIA classification algorithm to low quality data through fuzzy rankings and its application to the early diagnosis of dyslexia, *Neurocomputing,* 2016. DOI: 10.1016/j.neucom.2014.11.088.

[32] D. P. Pancho, J. M. Alonso, and L. Magdalena, Understanding the Inference Mechanism of FURIA by means of Fingrams, *Proc. 2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology,* 2015. DOI: 10.2991/ifsa-eusflat-15.2015.44.

[33] K. Trawiński, O. Cordón, and A. Quirin, A study on the use of multiobjective genetic algorithms for classifier selection in furia-based fuzzy multiclassifiers, *International Journal of Computational Intelligence Systems*, *5*, 2012, 231-253. DOI:: 10.1080/18756891.2012.685272.

[34] D. Tien Bui, T. C. Ho, B. Pradhan, B. T. Pham, V. H. Nhu, and I. Revhaug, GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and

MultiBoost ensemble frameworks, *Environmental Earth Sciences,75*(14), 2016. DOI: 10.1007/s12665-016-5919-4.

[35] V. R. Avali, G. F. Cooper, and V. Gopalakrishnan, Application of Bayesian logistic regression to mining biomedical data, *Proc. AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, 2014.

[36] Ioannis I. Spyroglou, Gunter Spöck, Eleni A. Chatzimichail, Alexandros Rigas and Emmanouil Paraskakis, A Bayesian Logistic Regression approach in Asthma Persistence Prediction, *Epidemiology Biostatistics and Public Health, 15*(1), 2018.

[37] Xia Zhao, Wei Chen, GIS-Based Evaluation of Landslide Susceptibility Models Using Certainty Factors and Functional Trees-Based Ensemble Techniques, *Appl. Sci., 10*(16), 2020,1-20. DOI:10.3390/app10010016

[38] MahmoodMoghadasian, SeyedehParvanehHosseini, Binary Cuckoo Optimization Algorithm for Feature Selection in High-Dimensional Datasets, *Proc. International conference on Innovative Engineering Technologies (ICIET'2014), Bangkok (Thailand),* Dec. 28-29, 2014, 18/21. DOI: org/10.15242/IIE.E1214027

[39] Sebastien C. Wong, Adam Gatt, Victor Stamatescu, Understanding data augmentation for classification: when to warp?, *Proc. 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, QLD, Australia,* 30 Nov.-2 Dec. 2016. DOI: 10.1109/DICTA.2016.7797091

[40] Yang, X.-S., & Deb, S., Cuckoo search: State-of-the-art and opportunities, *Proc. 2017 IEEE 4th International Conference on Soft Computing & Machine Intelligence (ISCMI),* 2017. DOI:10.1109/iscmi.2017.8279597

Biographies and Photographs

Samah Osama M. Kamel is Researcher in Dep. of Informatics at Electronic Research Institute, Egypt. Received B.S. degree in electronics and communications from Zagazig Faculty of Engineering, Zagazig University, in 2001. M.SC. "Secure IP Telephony Attack Sensor" at the Faculty of Engineering – Ph.D "Wireless Network Security System Analysis and Design" at the Faculty of Engineering–Shoubra, Electrical and Communication Engineering Department. There are many researches in network security and information security, machine learning and deep learning.

Sanaa Hassan AbouElhamayed is a PhD of engineering holder from Cairo, Egypt. She works as a part of Informatics Research Department in ERI. Her research interests are natural language processing, nformation system, and machine learning. She is a researcher in Electronic Research Institute. List of her latest publication: Enhancement of agriculture classification by using different classification systems. International Journal of Computer Applications (IJCA), 2016.Classifying datasets using some different classification methods.International Journal of Engineering and Technical Research (IJETR), 2016.Comparative Study on Different Classification Techniques for Spam Dataset.International Journal of Computer and Communication Engineering (IJCCE), 2016.