# An enhancement on support vector machine based on cuckoo algorithm

*Azadeh Jafari*
*Azad university of Kerman, Kerman, Iran*
***azadehjafari734@gmail.com***

**Abstract**

Classification is a highlighted area within data mining. Exploring data is the aim of data-mining to make understanding of a dataset. Among the various methods in data-mining, support vector machine (SVM) is well-known. The performance of SVM is highly depending on values of parameters, such as penalty factor and parameters in the Kernel part. Coco algorithm is an optimization algorithm that is used in non-linear problems, which has been proven to excel in other algorithms, such as genetic algorism (GA), particle swarm algorithms (PSO) and imperialist competitive algorithms (ICA) in terms of learning speed and convergence. This paper aims at applying coco-algorithm on breast cancer dataset and heart disease dataset.

**Keywords:** cuckoo algorithm, support vector machine (SVM), data-mining, classification

## 1- Introduction

Applying classification on medical data using developed coco-algorithm was, firstly, proposed in 2015 [6]. In 2014, using SVM and PSO in hybrid with coco algorithm were proposed; it consists of two stages: the first stage is based on (coco-search algorithms) CS to optimize the parameters of SVM, and the second stage PSO is applied on SVM to adjust its parameters. It was applied on two datasets, heart disease dataset and breast cancer. The results of CS-PSO-SVM showed its superiority, of classification accuracy, in compared to PSO-SVM and GA-SVM [3]. In 201, classification on long cancer using SVM kernel were proposed. In 2013, a method was proposed to recognize breast cancer using SVM classifier technique. During the past years, various features of breast tumors have been collected. Recognizing the cancer and filtering all information regarding the features of cancer recognition is a complex and time consuming issue. A dataset from WDBC (Diagnostic Wisconsin Breast Cancer) is taken to implement this method. This dataset were introduced in 1995, consisting of 32 features, where 6 out of 32 features are used to for training. The results not only show the recognition accuracy-rate of 98% but also decreasing the training time [4].

## 2- Data-mining

During past years, a large amount of data has been gathered in number of data warehouses. Due to vast varieties of stakeholders - audiences, consumers, markets, services, and etc. – and needs of accessing to proposer information for right and on-time decision making, it requires proper approaches to classify and explore effective information out of mass of data. Increase of amount of data, increase the interconnections between data, and thus, elevated the difficulty of exploring the meaning of data; where data-mining has great potentials to play the role. It is worth mentioning that abilities of data-mining for data analysis, extinguished the applications of conventional approaches [1].

## 3- Classification

Data-mining mostly relates to constructing models. A model, basically, is an algorithm or a rule set that maps a set of input to target. A model can make a suitable prediction within predefined circumstances. Data-mining consists of various algorithms that are deployed for every specific application; the most appealed one is the classification algorithm. The classification is a basic notion in areas of: statistics, machine learning, pattern recognition, and data-mining. There are various approached of data classification, e.g. multilayer perception artificial neural networks (MLP), radial basis function network (RBF), support vector machine (SVM), and K-nearest neighbors (KNN).

## 4- Artificial neural networks (ANN)

Leaning rule of for artificial neural networks were proposed in 1949; and later in 1958, MLP were proposed. It consists of three layers; which are: input layer, output layer and middle layer. Another approach is linear model to adjust neuron, which was proposed in 1960.

## 5- Support vector machine (SVM)

SVM were proposed in 1990 and established the theory of statistical leaning. SVM is usually applied for those problems with two classes in the output; however, there are proposed methods for those problems with multiple classes in the output by using hybridization of SVMs.

$$Lévy \sim u = t^{-}, \ (1<\lambda\leq 3) \qquad (1)$$

## 6- Cuckoo algorithm

Cuckoo optimization algorithm is an evolutionary algorithm that has been proposed after, a few widely used, algorithms such as genetic algorithms (GAs) and Simulated annealing (SA). It is in sequence of algorithms evolved from natural evolutions, e.g. particle swarm optimization (PSO), ant-colony optimization (ACO), artificial bee-colony (ABC), and artificial fish-swarm (AFS) [5]; where cuckoo algorithm is among recently proposed algorithm, in 2009, in gamut of meta-heuristic algorithms. To describe cuckoo search (CS) algorithm, as shown in Figure 1, we used three following rules: (i) every cuckoo lays only one egg at a time, and it keeps it in a random nest, (ii) the best nests with high quality of eggs will carry over to the next generation, and (iii) the number of host nests are fixed and the probability of discovering the egg by the host bird is $P_a \in (0,1)$. Each egg in nest represents a solution and every cuckoo give only one egg in a nest. This allows the chance of replacing better solutions in others nests with the current one. Obviously, this model is extendable from an egg to multiple eggs in a nest to represent multiple solutions. An advantage of CA over other algorithms, e.g. PSO and GA, is the simplicity. To produce new solutions, based on levy flights for cuckoos, we will have the equation as follows:

$$x_i^{(t+1)} = x_i^{t} + \propto \oplus Lévy(\lambda) \qquad (2)$$

Where, $\propto>0$ is a measurement for step scale; where we usually can use $\alpha=1$, and $\oplus$ shows the multiplication operation of input. The equation above is basically for random walk. This operation is similar to what PSO uses; however, the in CA, random walking through levy flight looks for better search domain.

$$Lévy \sim u = t^{-} , (1<\lambda\leq 3) \quad (3)$$

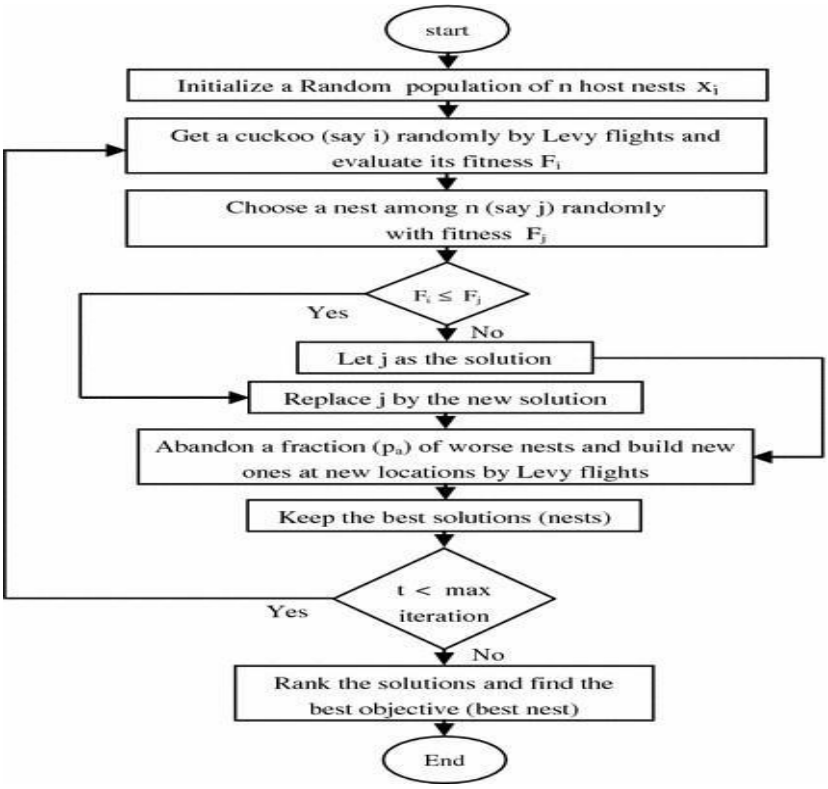


Figure 1: The flowchart of CS algorithm [7].

**7- Implementation**

The implementation of proposed method is done in MATLAB environment. Two datasets have been used for investigation, i.e. breast cancer dataset of Wisconsin, and heart disease of Cleveland. Table below indicates the details.

Table 1: description of used datasets

| | Cleveland | Wisconsin |
|---|---|---|
| Number of records | 270 | 699 |
| Number of records | 13 | 9 |
| Missing data | 11 | 16 |
| Type of data | Numerical and string | Numerical |
| Number of output | 2 | 2 |

To investigate the results of proposed method, the obtained results are compared with other two methods: (i) GA-based SVM, and (ii) ABC-based SVM. The number of iterations was set to 100 times, the population size to 50 solutions, and 10-fold cross validation. Figures 2 and 3, shows the accuracy of proposed method in comparison, on breast cancer and heart disease datasets consequently. Both figures show the superiority of using enhanced CA to increase the accuracy of SVM in compared with using GA or ABC.

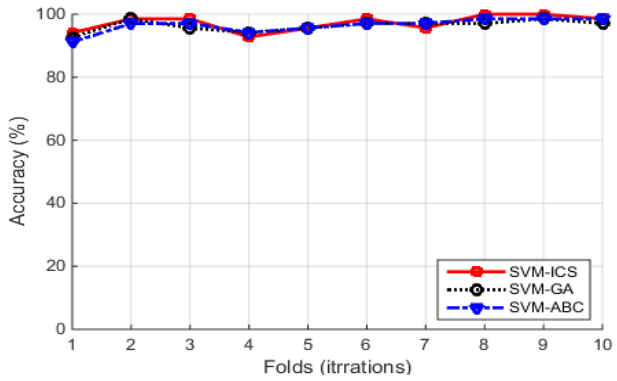


Figure 2: the accuracy of proposed method in comparison, on breast cancer dataset

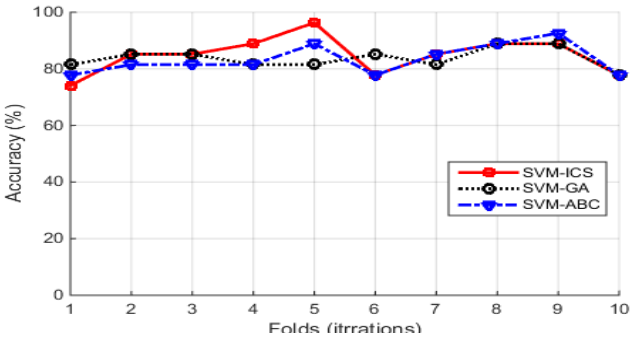


Figure 3: the accuracy of proposed method in comparison, on breast cancer dataset

## 8- Conclusion

This paper proposed a method based Cuckoo algorithms to hybrid with the SVM. The performance of proposed method was investigated in compared with equivalent GA-based and ABC-based methods. The implementation was done using MATLAB, and the results showed the superiority of proposed method when using Wisconsin breast cancer dataset and Cleveland heart disease dataset. It is remarkable that a kernel function must be chosen prior to train the algorithm, and then relate the optimized values to the algorithm parameters.

**References**

[1] Gaurav L. Agrawal, Prof. Hitesh Gupta, Optimization of C4.5 Decision Tree Algorithm for Data MiningApplication,March2013, Volume 3, Issue 3,341-345.

[2] Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni, Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction, March 2011, Volume 17– No.8.

[3] Zhiwen Liu, Hongrui Cao n, Xuefeng Chen, Zhengjia He, Zhongjie Shen, Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings 2014 , 399–410.

[4] Bichen Zheng, Sang Won Yoon , Sarah S. Lam, Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms, 2013 Elsevier Ltd. All rights reserved. 1476–1482.

[5] Bahman Deldadeh Barani , 2014, Support Vector Machine (SVM) Classifier.

[6] B. Jia, Biting Yu , Qi Wu, Xinshe Yang , Chuanfeng Wei , Rob Law , Shan Fu ,2016, Hybrid local diffusion maps and improved cuckoo search algorithm for multiclass dataset analysis, Neurocomputing 72.

[7] Bahman Deldadeh Barani , 2014,Support Vector Machine (SVM) Classifier.