Feature Selection using Artificial Bee Colony for Cardiovascular Disease Classification

B.Subanya, PG Scholar
Computer Science and Engineering
Kongu Engineering College,
Erode, India
subanyab@gmail.com

Dr.R.R.Rajalaxmi, Professor
Computer Science and Engineering
Kongu Engineering College,
Erode, India
rrr@kongu.ac.in

Abstract— **Machine learning techniques are widely used in medical decision support systems. Medical diagnosis helps to obtain different features representing the different variations of the disease. With the help of different diagnostic procedures, it is likely to have relevant, irrelevant and redundant features to represent a disease. Redundant features contribute to the wrong classification of the disease. Therefore, removing the redundant features reduces the size of the data and computation complexity. Identifying a good feature subset for effective classification is a non-trivial task. This requires an exhaustive search over the sample space of the dataset. The main objective of this paper is to use a metaheuristic algorithm to determine the optimal feature subset with improved classification accuracy in cardiovascular disease diagnosis. Swarm intelligence based Artificial Bee Colony (ABC) algorithm is used to find the best features in the disease identification. To evaluate the fitness of ABC, Support Vector Machine (SVM) classification is used. The performance of the proposed algorithm is validated against the Cleveland Heart disease dataset taken from the UCI machine learning repository. The experimental results show that, ABC–SVM performs better than Feature selection with reverse ranking. The results also show that, the proposed method obtained good classification accuracy with only seven features.**

Keywords—Feature Selection, Artificial Bee Colony, Optimization, Support Vector Machine

I. INTRODUCTION

Healthcare industry generates large amounts of complex data about patients, hospitals resources, disease diagnosis, electronic patient records, medical devices etc. The large amounts of data are a key resource to be processed and analyzed for knowledge extraction that enables support for cost–savings and decision making. Extracting useful knowledge and providing scientific decision–making for the diagnosis and treatment of disease from the database becomes necessary. Data mining in medical field brings a set of tools and techniques that can be applied to this processed data to discover hidden patterns that provide healthcare professionals an additional source of knowledge for making decisions. The datasets produced by biological research can be massive. The high dimensional nature of many modeling tasks in bioinformatics, going from sequence analysis over microarray analysis to spectral analyses has given rise to a wealth of feature selection techniques being presented in this field.

Cardiovascular disease also called as heart disease is a class of disease that involves the heart, the blood vessels or both. Major diagnostic involved are single photon emission computed tomography (SPECT) and electrocardiogram (ECG). A SPECT scan of the heart is a non-invasive nuclear imaging test. It uses radioactive tracers that are injected into the blood to produce pictures of heart. Doctors use SPECT images to diagnose coronary artery disease and find out occurrence of heart attack. ECG recording are analyzed to detect irregularity of heart beat or heart rhythm problems. As being easy to record, ECG is a first tool in diagnosis of heart diseases. The ECG provides significant clinical information of patients who have abnormal activity of heart. By using the ECG record physicians can classify the abnormality into which the disorder belongs. However, in the normal case the ECG is recorded in a long time period [10]. This is the much time–consuming process and inconvenient for the physician since he needs to be in alert at all times.

Feature selection is the method of selecting a feasible subset of features from the original set of candidate features. Unlike feature extraction, feature selection method is applied to datasets with known features. These methods will attempt to identify the significant features and discard irrelevant or redundant features from the original set of features. Feature selection methods can be classified into three major categories based on the technique of the search and selection process: complete, stochastic and heuristic search.

The main objective of this paper is to use a metaheuristic algorithm to determine the optimal feature subset with improved classification accuracy in cardiovascular disease diagnosis. Swarm intelligence based Artificial Bee Colony (ABC) algorithm is used to find the best features in the disease identification. To evaluate the fitness of ABC, Support Vector Machine (SVM) classification is used.

The rest of the paper is organized as follows: Section 2 provides an overview of feature selection methods. Section 3 reviews the relevant literature on feature selection, artificial bee colony and SVM. Section 4 discusses proposed artificial bee colony algorithm (ABC). Section 5 then discusses the experimental procedure used with the dataset. Experimental results are compared with those of existing approaches in section 6. Conclusions are finally drawn in Section 7.

II. FEATURE SELECTION METHODS

A typical feature selection process consists of four basic steps namely, subset generation, subset evaluation, stopping criterion and result validation. Besides the reduction in the number of features, the accuracy of the selected feature subset is critically important in feature selection. Thus, it is used in conjunction with classification techniques to model and learn the underlying processes in bioinformatics. There are several approaches for combining the feature selection with the classification methods. The three common methods of feature selection used are filter technique, embedded technique and wrapper technique.

A. Filter Method

The filter model relies on general characteristics of the data to evaluate and select feature subsets without involving any mining algorithm. The operations of feature selection and classification are performed separately and independently from one another. Many feature selection methods adopt statistical functions to evaluate the quality of each feature in the feature selection process. The output is the optimal subset of features without redundancy or noise. These selected features are then presented as input data for classification algorithms and accuracy is evaluated. This technique is fast and scalable to most of the high–dimensional problems, but it has poor classification performance.

B. Embedded Method

Embedded techniques perform feature selection as part of the learning procedure and are usually specific to given learning machines. The search for an optimal subset of features is built into the classifier construction, and can be seen as a search in the combined space of feature subset and hypotheses. Examples are classification trees, random forests, feature selection using weight vector of Support Vector Machines (SVM) and methods based on regularization techniques. These methods are less computationally intensive than wrapper methods. The embedded feature selection techniques must be designed specifically for a learning algorithm. However, it is very complex to construct a mathematical model for a feature selection embedded classifier.

C. Wrapper Method

The feature selection is wrapped around instead of embedding into a classifier. The feature selection methods use accuracy from the classifier to select the feasible subset of features. Therefore, different classification techniques and different selection procedures will produce different sets of selected features. This method requires more computation time than the embedded techniques. However it is simple to implement and interacts with the classification method whereas the filter methods are independent of the classification methods.

III. LITERATURE REVIEW

Data overload is a serious problem, because the data analysis requires more computational resources and consumes much time. Hence, feature selection process is used to remove the irrelevant or noise features from the data [9] in order to reduce the time and the resource usage. The reduction in the number of features should not reduce the accuracy in finding the disease. So, feature selection is used in combination with the classification techniques. The following sections discuss the several feature selection methods and classification processes.

Feature selection algorithms used for classification and clustering are discussed in [7]. The authors grouped and compared different algorithms with a categorizing framework based on search strategies, evaluation criteria, data mining tasks, and provided guidelines in selecting feature selection algorithms. They also presented an illustrative example as to how the existing feature selection algorithms can be integrated into a meta algorithm that takes advantage of individual algorithms. They proposed two architectures: a categorizing framework and a unifying platform. The unifying platform helps to integrate various methods. Real–world application of feature selection also has been given and it helps any new person in the verge of feature selection to choose the algorithms that suit for that particular application and data mining tasks.

An algorithm for arrhythmia classification, which is associated with the reduction of feature dimensions by Linear Discriminant Analysis (LDA) and a SVM based classifier was proposed [8]. Seventeen original input features were extracted from preprocessed signals by wavelet transform, and then those features were reduced to four features by LDA. The performance of the SVM classifier with reduced features by LDA is better than Principal Component Analysis (PCA) and even with original set of features. For a cross–validation procedure, this SVM classifier was compared with Multilayer Perceptron (MLP) and Fuzzy Inference System (FIS) classifiers. While all classifiers used the same reduced features, the overall performance of the SVM classifier was higher than others.

Simulated Annealing (SA) approach [6] was applied for parameter determination and feature selection in SA–SVM approach. The biggest difficulties in setting up the SVM model are choosing the kernel function and its parameter values. If the parameter values are not set properly, then the classification outcomes will be less than optimal. Wrapper approach uses accuracy as the parameter for evaluation. A classifier is constructed with the aim of maximizing the predictive accuracy. The features utilized by the classifier are then selected as the optimal features. The results of SA–SVM approach without feature selection were compared with the SVM to test several datasets from University of California, Irvine (UCI). Results clearly indicate that SA-SVM was much helpful in selecting the parameter values.

Linear Forward Selection technique helped to reduce the number of attribute expansions in each forward selection

step[4]. The experiments demonstrated that this approach is faster and finds smaller subsets and can even increase the accuracy compared to standard forward selection. The results showed that this approach leads to competitive results, requires less runtime, and has less over fitting compared to complete forward selection. Many works have been carried out in the past to speed up the wrapper. Two variants of linear forward selection such as fixed width and fixed–set width have been presented. They are preferable to standard forward selection, mainly because of the dramatic reduction in runtime, but also because they can produce smaller subsets without much reduction in accuracy.

An improved metaheuristic based on Greedy Randomized Adaptive Search Procedure (GRASP) for the problem of feature selection has been developed [3]. GRASP–FS approach provides an effective scheme for wrapper–filter hybridization. Five benchmark datasets available in UCI repository were used to validate GRASP components like Sonar, Ionosphere, SpamBase, Audiology and Arrhythmia with respectively 60, 34, 57, 69 and 279 attributes. They investigated the GRASP component design as well as its adaptation to feature selection problem. Results confirmed the robustness of the hybridization schemata.

A hybrid filter–wrapper feature subset selection algorithm based on Particle Swarm Optimization (PSO) for SVM classification was built for feature selection [2]. The filter model is based on the mutual information and is a measure of feature relevance and redundancy with respect to the feature subset selected. The wrapper model is a modified discrete PSO algorithm. This hybrid algorithm is named as maximum relevance minimum redundancy PSO (mr^2PSO). The performance of the proposed algorithm has also been compared with hybrid filter–wrapper algorithm based on a genetic algorithm and a wrapper algorithm based on PSO. The results showed that the mr^2PSO algorithm is competitive in terms of both classification accuracy and computational performance.

Medical knowledge driven Feature Selection (MFS) along with the generally employed computational intelligence [5] based feature selection mechanism helped to select important features. MFS combined with the Computerized Feature Selection process (CFS) has also been investigated and showed encouraging results particularly for Naive Bayes and Sequential Minimal Optimization (SMO). Waikato Environment for Knowledge Analysis (WEKA) has been used for implementation and they tested with UCI Heart Disease dataset. In order to provide a comparison among the well popular classification algorithms, they have considered four performance metrics. They are accuracy, true positive rate (TP), F–measure, and time. Here, accuracy was the overall prediction accuracy, true positive rate (TP) was the accurate classification rate for the positive classes, and F–measure indicates the effectiveness of an algorithm when the accurate prediction rates for both of the classes are considered. Also, training time was considered to compare the computational complexity for learning. Feature selection based on medical knowledge is an important factor in heart disease diagnosis. If significant symptoms related to heart disease are not considered then there is a strong likelihood that the diagnosis neglects the important factors.

A novel feature selection approach for the classification of high dimensional cancer microarray data using filtering signal-to-noise ratio (SNR) score and Particle swarm Optimization (PSO) was proposed in [10]. At first the data set is clustered using k-means clustering and SNR score is used to rank genes. High scored genes are gathered from each cluster to form a new feature subset. Secondly, the generated feature subset is given to PSO and the optimized feature subset is produced. For evaluating the feature subsets, Support vector machine (SVM), k-nearest neighbor (k-NN) and Probabilistic Neural Network (PNN) classification methods with leave one out cross validation approach is used. The results illustrate that the proposed approach using PSO gives better result than others.

ABC was proposed as a method for data dimension reduction in classification problems [11]. It is used to select the optimal subset of dimensions from the original high–dimensional data. The k–Nearest Neighbor (k–NN) method is then used for fitness evaluation within the ABC framework. ABC and k–NN have been modified and bundled together to create an effective dimension reduction method. ABC uses the behaviour of three types of bees namely employed bees, onlooker bees, and scout bees. Proposed method applies ABC wrapping with a k–NN classifier. k–NN is used for fitness evaluation to estimate the fitness value of the ABC food sources. After the employed bees and the onlooker bees generate new candidate food sources, which are a subset of selected features, k–NN is performed to evaluate the classification accuracy of the new candidate food sources. The accuracy is used as a criterion for selecting the optimal subset of features. The existing diagnostic tools are based on interview and behaviour observation, which is extremely time-consuming. They carried out the experiments with gene expression data and autism dataset. The results of the gene expression analysis showed that the ABC – k–NN method can effectively reduce the data dimension while maintaining high classification accuracy.

Modified Binary Particle Swarm Optimization (MBPSO) method for feature selection with the simultaneous optimization of SVM kernel parameter setting helped to optimize feature selection [12]. They applied that to mortality prediction in septic patients. An enhanced version of Binary Particle Swarm Optimization (BPSO) has been designed to avoid the premature convergence of the BPSO algorithm is proposed. MBPSO control the swarm variability using the velocity and the similarity between best swarm solutions. This method uses SVM in a wrapper approach, where the kernel parameters are optimized at the same time. The generation of models for the mortality risk evaluation in patients with sepsis is important topic in medicine. The approach has been applied to predict the outcome (survived or deceased) of patients with septic shock. MBPSO has been tested with several benchmark datasets and is compared with other PSO based algorithms and Genetic Algorithms (GA). The experimental results showed

that the proposed approach can correctly select the discriminating input features and also achieve high classification accuracy, when compared to other PSO based algorithms. When compared to GA, MBPSO is similar in terms of accuracy, but the subset solutions have less selected features.

A hybrid model which combines filter and wrapper approach to achieve better classification performance for cardiovascular disease diagnosis has been proposed in [10]. Three algorithms are implemented with hybrid model and SVM classifier. They are Forward Feature Inclusion, Back–elimination Feature Selection and Forward Feature Selection. The features are ranked using distance criterion and then wrapper model is used to evaluate classification model. These criterions generate all features ranks as per their importance towards target class identification.

Datasets produced by biological process can be massive. Some research needs more factors or features that need to be analysed in diagnosing a particular disease. But high dimensionality of the feature space affects the classification accuracies and the computational complexity due to redundant, irrelevant and noisy features present in the dataset. One possible solution is to use the feature reduction techniques. Many feature selection methods have been proposed. Comparisons have been made between them. Filter, wrapper, embedded approaches are used in combinations. Based on the survey made wrapper approach yielded highest classification accuracy and were widely used in biological research. In order to improve the classification accuracy in predicting the heart disease a wrapper technique is proposed.

IV. ARTIFICIAL BEE COLONY ALGORITHM

In the proposed feature selection approach, ABC algorithm optimizes the process of feature selection and yields the best optimal feature subset which increases the predictive accuracy of the classifier. ABC is used as a feature selector and generates the feature subsets and a classifier is used to evaluate each feature subset produced by the onlookers. Hence the proposed system is a wrapper based system. The following steps illustrate the proposed algorithm:

1. Cycle =1
2. Initialize ABC parameters
3. Evaluate the fitness of each individual feature
4. Repeat
5. Construct solutions by the employed bees
   - Assign feature subset configurations (binary bit string) to each employed bee
   - Produce new feature subsets
     $$\upsilon_{ij} = x_{ij} + \Phi_{ij}\left(x_{ij} - x_{kj}\right)$$
   - Pass the produced feature subset to SVM classifier
6. Construct solutions by the onlookers
   - Select a feature based on the probability P_i
   - Compute v_i using x_i and x_j
   - Apply greedy selection between v_i and x_i
7. Determine the scout bee and the abandoned solution
8. Calculate the best feature subset of the cycle
9. Memorize the best optimal feature subset
10. Cycle = Cycle + 1
11. until pre–determined number of cycles is reached
12. Employ the same searching procedure of bees to generate the optimal feature subset

A. *Classification Using Support Vector Machines*

SVM is widely used [1] in computational biology due to their high accuracy, their ability to deal with high–dimensional and large datasets, and their flexibility in modeling diverse sources of data. The simplest form of a prediction problem is binary classification: trying to discriminate between objects that belong to one of two categories: positive (+1) or negative (–1) SVMs use two key concepts to solve this problem: large margin separation and kernel functions. The idea of large margin separation can be motivated by classification of points in two dimensions. A simple way to classify the points is to draw a straight line and call points lying on one side as positive and on the other side as negative. Kernels are used for nonlinear classification. Many real life problems require classification of more than two classes.

SVM performs classification by constructing an n–dimensional hyper plane that optimally separates the data into two categories.

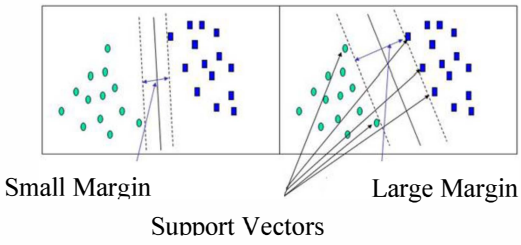

Figure 1 SVM Hyper Plane and Support Vectors

A good separation is achieved by the hyper plane that has the largest distance to the nearest data points in the training set (maximum margin). Figure 1 shows the SVM hyper plane and support vectors. SVM can efficiently perform for non–linear classification with the help of kernel function. Some common kernels are

- Linear Kernel
- Polynomial kernel
- Radial basis function (RBF) kernel
- Sigmoid kernel

V. EXPERIMENTAL PROCEDURE

The proposed model is implemented using MATLAB running on Intel core i3 processor with a 4 GB RAM capacity and a 320 GB hard disk. Cleveland dataset is taken from UCI

repository. The attributes of the Cleveland dataset with their description are given in Table I.

TABLE I. ATTRIBUTE INFORMATION

S.No.	Feature	Description
1	Age	Age in years
2	Sex	Male, female
3	Cp	Chest pain type Value 1: typical angina Value 2: atypical angina Value 3:non–anginal pain Value 4: asymptomatic
4	Trestbps	Patient's resting blood pressure in mm Hg at the time of admission to the hospital
5	Chol	Serum cholesterol in mg/dl
6	Fbs	Boolean measure indicating whether fasting blood sugar is greater than 120 mg/dl (1 = True; 0 = false)
7	Restecg	Electrocardiographic results during rest
8	Thalach	Maximum heart rate attained
9	Exang	Boolean measure indicating whether exercise induced angina has occurred
10	Oldpeak	ST depression brought about by exercise relative to rest
11	Slope	The slope of the ST segment for peak exercise
12	Ca	Number of major vessels (0–3) coloured by fluoroscopy
13	Thal	The heart status (normal, fixed defect, reversible defect)
14	Class	Class attributes

A. PARAMETERS FOR EVALUATION

Confusion matrix is useful for analyzing the performance of a classifier and it is shown in Table II. TP and TN tell us when the classifier is getting things right, while FP and FN tell us when the classifier is getting things wrong. Using training data to derive a classifier and then estimating the accuracy of the resulting learned model can lead in misleading estimates due to over–specialization of the learning algorithm. Instead, it is better to measure the classifier's accuracy on a test set consisting of class–labeled tuples that were not used to train the model.

TABLE II. CONFUSION MATRIX

		PREDICTION	
		0	1
ACTUAL	0	TN	FP
	1	FN	TP

A population of tested individuals for the presence of heart disease may be divided into four groups,

- True Positives: those who test positive for a condition and are positive (TP)
- False Positives: those who test positive, but are negative (FP)
- True Negatives: those who test negative and are negative (TN)
- False Negatives: those who test negative, but are positive (FN)

Classifier Accuracy or recognition rate denotes the percentage of test set tuples that are correctly classified. It is given by the formula,

$$Accuracy = \frac{TP + TN}{TOTAL\_INPUT}$$

VI. RESULTS AND DISCUSSION

In order to evaluate the performance of the proposed method, the following control parameters are used.

- Number of dimensions =14
- Number of employed bees and onlooker bees =12
- Maximum number of iterations =100
- Limit=10

Table III gives the comparison of result between the forward feature selection and the proposed ABC–SVM method. The results show that, ABC–SVM performs better than Feature selection with reverse ranking. The results also show that, the proposed method obtained good classification accuracy with only seven features.

TABLE III. COMPARISON BETWEEN FORWARD FEATURE SELECTION AND ABC–SVM METHOD

Forward Feature Selection			ABC–SVM	
Method	**No. of Features Obtained**	**Accuracy**	**No. of Features Obtained**	**Accuracy**
Forward ranking	4	85 %	7	86.76 %
Reverse ranking	8	85 %		

Table IV gives about the details of the features obtained by the ABC–SVM method. In the first experiment, ABC–SVM yielded an accuracy of 85.29% with five features. For the

same dataset, in the second experiment it produced an accuracy of 86.76 % with seven features. It is found that, Age and fasting blood sugar are the prominent features in cardiovascular disease.

TABLE IV. FEATURES OBTAINED WITH ABC–SVM

Feature Name	Accuracy
Chol, Restecg, Thalach, Oldpeak, Slope	85.29 %
Age, Chol, Fbs, Restecg, Thalach, slope, ca	86.76 %

VII. CONCLUSION

Feature selection is mainly used for selecting the best attributes from the given data especially in medical diagnosis. Heuristic methods help to resolve the problem of selecting best features. ABC is a metaheuristic algorithm that share information between the bees in the population and select feasible solutions, which can satisfy the defined criteria. ABC has a unique solution update mechanism (updation in two phases), which allows the results to converge to the optimal solution quickly. Also, it is simple and easy to implement because it has fewer control parameters to configure.

In this paper, ABC-SVM used the wrapper technique for classification and the experimental results of the algorithm with the heart disease data showed improvement in accuracy when comparing to the conventional forward selection and back elimination feature selection. The algorithm identified seven features for disease identification.

References

[1] Aksu Y., Miller D.J. and Kesidis G. (2010), 'Margin-maximizing feature elimination methods for linear and nonlinear kernel-based discriminant functions', IEEE Trans. Neural networks, pp.701-717

[2] Alper U., Alper M. and Ratna B.C. (2011), 'mr2PSO:A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification', Information Science 181, pp.4625-4641

[3] Esseghir M.A. (2010), 'Effective Wrapper-Filter hybridization through GRASP schemata JMLR', In the fourth workshop on feature selection in data mining, workshop on feature selection in data mining, workshop and conference proceedings 10, pp.45-54

[4] Gutlein M., Frank E. and Karwath A. (2010), 'Large-scale attribute selection using wrappers', In IEEE symposium computational intelligence and data mining, pp.332-339

[5] Nahar J.T.I, Kevin S.T. and Y.P.chen. (2013), 'Computational Intelligence for heart disease diagnosis: A medical knowledge driven approach', Expert systems with applications 40, pp.96-104

[6] Lin S.W., Lee Z.J., Chen S.C and Tseng T.Y. (2008), 'Parameter determination of support vector machine and feature selection using simulated annealing approach', Appl. Soft computing 8, pp.1505–1512

[7] Liu H. and Yu L. (2005), 'Toward integrating feature selection algorithms for classification', IEEE Trans. Knowledge Data Eng

[8] Song.M.H., Lee.J, Cho.S.P., Lee.K.J and Yoo.S.K (2005), 'Support vector machine Based Arrhythmia classification using reduced Features', International Journal of control, Automation, and Systems, Vol 3, pp.571-579

[9] Sahua B. and Mishra D. (2012), 'A Novel Feature Selection algorithm using Particle Swarm Optimization for cancer Microarray Data', Procedia Engineering 38, pp.27-31

[10] Swati S. and Ashok G. (2013), 'Feature selection for medical diagnosis: Evaluation for cardiovascular diseases', Expert Systems with applications 40, pp.4146-4153

[11] Thananan Prasartvit., Anan Banharnsakun., Boonserm Kaewkamnerdpong. and Tiranee Achalakul. (2013), 'Reducing bio-informatics data dimension with ABC-KNN', Neurocomputing 116, pp. 367-381

[12] Vieira.S.M, Luis F. Mendonca.L.F, Farinha.G.J. and Sousa M.C.J (2013), 'Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients', Applied soft computing 13, pp.3494-3504.