

Artificial Bee Colony based Feature Selection for Effective Cardiovascular Disease Diagnosis

Subanya B, R.R. Rajalaxmi

Abstract— Machine learning has been an effective support system in medical diagnosis which involve large amount of data. Analyzing such data consumes more time in terms of execution and resources. All data features do not support for the end results. Hence it is very important to identify the features that contribute more in identifying the diseases. Those with less contribution can be eliminated. The need of feature selection arises when we need to reduce the massive medical data to reduced number of features. The objective of this paper is to design an effective algorithm that can remove irrelevant dimensions from large data and to predict more accurately the presence of disease. Artificial Bee Colony based feature selection is incorporated and a wrapper classifier is used for classification. A Binary Artificial Bee Colony (BABC) algorithm is used to find the best features in the disease identification. The fitness of BABC is evaluated using Naive Bayesian method. Results are validated using Cleveland Heart disease dataset taken from the UCI machine learning repository. The results indicate that, BABC–Naive Bayesian outperform the other methods.

Index Terms— Feature Selection, Binary Artificial Bee Colony, Optimization, Naive Bayesian method, Cardiovascular Diseases

1 INTRODUCTION

In today's competitive world, humans are affected by different kinds of life-threatening diseases. Out of this, heart disease has received more attention. Heart disease refers to the injury of the heart and blood vessels. Some common symptoms [Health, 2010] of this disease are chest pain, discomfort in the chest area, shortness of breathe, sweating, dizziness, palpitation etc., However, the above symptoms differ from person to person. The main causes of heart disease found to be age, family history, eating habits, blood pressure, diabetes, gender, life style etc., Medical diagnosis helps to identify the symptoms and causes for this disease. The diagnostic procedures produce information about the different variations of the disease.

Data mining explore hidden patterns from large volume of data. Extracting useful knowledge and providing scientific decision-making for the diagnosis and treatment of disease from the database becomes necessary. Different data mining techniques can be applied to the heart diagnosis data to provide healthcare professionals an additional source of knowledge for making decisions. The datasets produced by different diagnostic procedures can be massive. The high dimensional nature of the data has given rise to a wealth of feature selection techniques being presented in this field.

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.

This paper employs a metaheuristic algorithm to determine the optimal feature subset with good classification accuracy for heart disease diagnosis. BABC algorithm is used to find the best features for disease classification. The selected feature subsets are validated using Naive Bayesian method of classification.

The rest of the paper is organized as follows: Section 2 presents a brief introduction of feature selection methods. Section 3 provides literature survey of feature selection, artificial bee colony and Naive Bayesian method. Section 4 specifies proposed binary artificial bee colony algorithm. Section 5 then discusses the experimental procedure used with the dataset. Experimental results are compared with those of existing approaches and the performance analysis is done in section 6. Conclusions are finally drawn in Section 7.

2 FEATURE SELECTION METHODS

Typically, feature selection process contains the following methods: subset generation, subset evaluation, stopping criterion and result validation. Eventhough, the aim of this process is to remove irrelevant and redundant features, the generated subset accuracy is more important. Thus it is essential to validate it using the classification techniques to model and learn the underlying processes in bioinformatics. Researchers proposed various methods to combine feature selection with the classification methods. The three common methods of feature selection used are filter technique, embedded technique and wrapper technique.

- Subanya B is currently pursuing masters degree program in Computer Science and Engineering in Kongu Engineering College, Perundurai, India PH-+919566022577. E-mail: subanyab@gmail.com
- Dr. R. R. Rajalaxmi is currently working as a professor in Computer Science and Engineering department in Kongu Engineering College, Perundurai, India PH-+919486561199. E-mail: rrr@kongu.ac.in

2.1 Filter Method

This model uses general characteristics of the data to evaluate and select feature subsets without including any mining algorithm. Feature selection and classification are performed independently from one another. In this category, many feature selection methods use 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. Based on the selected features, the accuracy is computed with a classification algorithm. Since, this method is independent from the classification task, it is more suitable for high dimensional data. But, it has poor classification performance.

2.2 Embedded Method

Feature selection is a part of the learning procedure and specific to given learning machines. Here, optimal feature subset selection is built into the classifier construction, and viewed as a search in the combined space of feature subset and hypotheses. Some examples of this method include, 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 design of embedded feature selection techniques depend on a specific a learning algorithm. But, it is very complex to construct a mathematical model for a feature selection embedded classifier.

2.3 Wrapper Method

This method uses wrapper approach for feature selection instead of embedding into a classifier. The selected feature subsets are verified with the help of classification algorithms features. Hence, it is possible to get different subsets based on different classification techniques. But, the computational complexity is high in comparison to embedded and filter methods. As it is simple to implement and interacts with the classification method, more work is carried using this method than filter and embedded methods.

3 LITERATURE REVIEW

The data analysis requires more computational resources and consumes much time when that data is of huge volume. Hence, feature selection process is used to remove the irrelevant or noise features from the data [12] 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 [9]. 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 [11]. 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 [8] 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 pro-

vides an effective scheme for wrapper-filter hybridization. Five benchmark datasets available in UCI repository were used to validate GRASP components like Sonar, Ionosphere, Spam-Base, 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 schema.

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 (mr²PSO). 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 mr²PSO 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 [7] 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 hybrid neural network that includes artificial neural network (ANN) and fuzzy neural network (FNN) has been developed [6]. The main aim of the paper is to classify the data of the heart and diabetes diseases better so as to have more reliable diagnosis. Fuzzy and crisp values are used as together in medical data. Two real-time problem data were investigated for determining the applicability of the proposed method. The data were obtained from the UCI machine learning repository. The datasets are Pima Indians diabetes and Cleveland heart disease. In order to evaluate the performance of the proposed method accuracy, sensitivity and specificity performance measures that are used commonly in medical classifica-

tion studies were used. The proposed method achieved accuracy values 84.24% and 86.8% for Pima Indians diabetes dataset and Cleveland heart disease dataset respectively. The classification accuracies of these datasets were obtained by k-fold cross-validation.

A novel feature selection approach for the classification of high dimensional cancer microarray data using filtering the signal-to-noise ratio (SNR) score and Particle swarm Optimization (PSO) was proposed in [11]. 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, SVM, 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 [16]. It is used to select the optimal subset from the original high-dimensional data. 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 [17]. 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 [14]. 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 criteria generate all features ranks as per their importance towards target class identification.

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 (update in two phases) which allows the results to converge to the optimal solution quickly. Swarm intelligence based Artificial Bee Colony (ABC) algorithm has been proposed to find the best features in the disease identification. To evaluate the fitness of ABC, SVM [13] classification is used. The performance of the proposed algorithm is validated against the Cleveland Heart disease dataset taken from the UCI machine learning repository with the 303 samples. The experimental results showed that ABC-SVM performs better than forward feature selection with reverse ranking. The results also showed that the proposed method obtained good classification accuracy with only seven features.

Datasets produced by biological process can be massive. Some research needs more factors or features that need to be analyzed 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.

4. BINARY ARTIFICIAL BEE COLONY ALGORITHM

In real-coded Artificial Bee Colony (ABC) algorithm the bees move around the search space within the continuous and real domain. Three types of bees namely employed bees, onlooker

bees and scout bees are involved in the process. Since feature selection deals with search in binary domain it is essential to adopt discrete space while updating the candidate and neighborhood solutions. Thus, the solution for feature selection is represented as shown in Table I.

TABLE I: SOLUTION REPRESENTATION

Objects	Features				
	f_1	f_2	f_3	f_n
O_1	1	0	1	1
O_2	0	1	1	0
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
O_m	0	0	1		1

Let n denote the number of features associated with m objects. If the feature j is selected for an object i , then "1" is assigned. Otherwise, "0" is assigned. This encoding represents the food source as a matrix of size $m \times n$. The Binary ABC (BABC) consists of the following steps:

1. Set the initial values for number of employed bees, number of onlooker bees, maximum number of iterations, and pre-determined number of iterations (*limit*) for scout bees.
2. Initialize the population of food sources (possible set of features) randomly and assign it to the employed bees using the binary solution representation as shown in Table I.
3. Evaluate the initial population of food sources using the following objective function :

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TOTAL_INPUT}} \quad (1)$$

Compute accuracy using k-NN classifier.

4. *Employed bees*: Generate new candidate solutions generated in the neighborhood of existing solutions by swapping any two randomly selected objects (whole columns in the matrix) in the current solutions (one for each employed bee). The new subset is formed using the formula

$$v_{ij} = x_{ij} + \Phi_{ij} (x_{ij} - x_{kj}) \quad (2)$$

If random selection picks the same object, then repeat the process. After that, evaluate the solutions using k-NN classifier.

5. *Update Employed bees*: For each employed bee, if its candi-

date solution (food source) is better than its present one, then replace it with the candidate solution.

6. *Onlooker bees*: For the onlooker bees, select one of the employer bee food sources probabilistically based on the following equation.

$$P_i = \frac{\text{accuracy}_i}{\sum_{k=1}^n \text{accuracy}_j} \quad (3)$$

7. *Onlooker bee* Generate new solution for each onlooker then by swapping any two randomly selected columns of their selected food sources. If the new food source fitness is higher than the existing food source, then update the corresponding employed bee's position to the new food source. Otherwise, do not change the employed bee food source.
8. *Scout bees*: Abandon the food source of employed bee if the current food source fitness is not improved within pre-determined number of iterations (Limit) and make the corresponding as a scout bee and generate a new food source (solution) randomly.
9. Memorize the best solution obtained so far and repeat from step 3 until a predetermined termination criterion is met.

4.1 Naive Bayesian Classifier

The Bayesian Classification represents a supervised learning method as well as a statistical method for classification. Naive Bayesian Classification is based on the Bayesian theorem and is particularly suited when the dimensionality of the inputs is high. Parameter estimation for naive Bayes models uses the method of maximum likelihood. It performs better in many complex realworld situations and the advantage is that it requires a small amount of training data to estimate the parameters.

Working of Naive Bayesian classification is as follows. There are classes, say C_k for the data to be classified into. Each class has a probability $P(C_k)$ that represents the prior probability of classifying an attribute into C_k . The values of $P(C_k)$ can be estimated from the training dataset. For n attribute values, v_j , the goal of classification is clearly to find the conditional probability $P(C_k | V_1 \wedge V_2 \wedge \dots \wedge V_n)$. By Bayes' rule, this probability is equivalent to

$$\frac{P(C_k | V_1 \wedge V_2 \wedge \dots \wedge V_n) P(C_k)}{P(V_1 \wedge V_2 \wedge \dots \wedge V_n)}$$

For classification, the denominator is irrelevant, since, for given values of the v_j , it is the same regardless of the value of C_k . The central assumption of Naive Bayesian Classification is that within each class the values V_j are all independent of each other. Then by the laws of independent probability,

$$P(V_i | \{\text{all the other values of } V_j\}, C_k) = P(V_i | C_k)$$

and therefore

$$P(V_1 \wedge V_2 \wedge \dots \wedge V_n | C_k) = P(V_1 | C_k) P(V_2 | C_k) \dots P(V_n | C_k)$$

Each factor on the right-hand side of this equation can be determined from the training data, because (for an arbitrary V_i),

$$P(V_i | C_k) \approx [\#(V_i \wedge C_k)] / [\#(C_k)]$$

where “#” represents the number of such occurrences in the training set data. Therefore, the classification of the test set can now be estimated by

$P(C_k | V_1 \wedge V_2 \wedge \dots \wedge V_n)$ which is proportional to

$$P(C_k) P(V_1 | C_k) P(V_2 | C_k) \dots P(V_n | C_k)$$

4.2 Fitness Function I

A fitness function is needed to guide the search by assigning to any tentative solution a quality value. Once the accuracy value and the number of genes are known, the fitness function is calculated according to :

$$f(x) = (100 - \text{acc}) + \lambda \frac{\#(\text{features in subset})}{\#(\text{total feature})}, \text{ being } \lambda = 10^{\log(\#(\text{total feature})) + 1} \quad (4)$$

The objective here consists of maximizing the accuracy and minimizing the number of genes. For Convenience (only maximizing of fitness), the first factor presented as $(100 - \text{acc})$ and the second one is normalized in order to control the tradeoff between these two factors.

A constant value λ (which depends on the total number of features) is used in this normalization. Therefore if the number of features in the subset is high (with regards to the total numbers of features in the original dataset), then the fitness function promotes the reduction of features. Otherwise, if the number of features in the subset is small, then this fitness function promotes the improvement in accuracy.

Initially the experiment is carried out with accuracy alone as the fitness function. Equation (1) is used to calculate the accuracy and the same experiment is carried out by changing the fitness function and is calculated from (4). The accuracy taken from the Naive Bayesian classifier is fed in to acc part of (4) and the experiments have been carried out.

5. DATASET DESCRIPTION

This work used the most popular and publically available Cleveland heart disease dataset from UCI machine learning repository. Eventhough, it contains 76 attributes majority of the work utilized 14 attributes. 14 attributes were taken as relevant attributes in identifying the heart disease data. The total number of samples present in the dataset is 303. The attributes of the Cleveland dataset with their description are given in Table II.

TABLE II. 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

Many real world datasets contain missing values due to data entry mistakes, incorrect measurements, equipment malfunction and lack of verification from the sources. While analyzing the dataset, six records contain missing values. Since presence of missing values lead to degrade the performance of classification, the records containing missing values can be removed. On the other hand data imputation methods help to compute the approximate probable missing values. One such method is mean imputation. Here, the missing values of an attribute are replaced by the mean value of the attribute.

6. PERFORMANCE ANALYSIS

Accuracy is the most common performance metric used in medical diagnosis tasks. It is calculated as follows:

$$\text{Accuracy} = \frac{\text{Total no. of correctly diagnosis cases}}{\text{Total no. of cases}} \quad (5)$$

In order to evaluate the performance of the BABC Naive Bayesian algorithm, we validated the proposed algorithm in comparison to other algorithms used for heart disease diagnosis [6]. In the experiment, the parameter settings of the proposed method are as follows:

- Number of dimensions =13

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

Table III gives the comparison of result between the BABC Naive Bayesian and other methods. ABC-SVM [13] 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. BABC- Naive Bayesian produced 86.4 % accuracy with six features. The results showed that, BABC Naive Bayesian performs better than all the methods in terms of accuracy. The algorithm produced good accuracy with six features namely Cp, Trestbps, Chol, Thalch, Slope and Thal.

TABLE III. COMPARISON BETWEEN BABC NAIVE BAYESIAN WITH OTHER METHODS

Method	Accuracy	No. of features
BABC- Naive Bayesian	86.4%	6
BABC- Naive Bayesian (Fitness function I)	85.5 %	7
ABC-SVM	86.76 %	7
IncNet	90%	All
Hybrid Neural Network	86.8 %	All
LDA	84.2%	All
Naive Bayes	83.1%	All
IBPSO k-NN	82.15%	10
SVM (5 × CV)	81.5%	All

7. CONCLUSION

Feature selection in medical diagnosis helps to improve clinical decision making. Heuristic methods help to resolve the problem of selecting best features. In this work, a metaheuristic algorithm BABC-Naive Bayesian is developed to determine the best features in heart disease diagnosis. The experimental results indicate that the proposed algorithm can effectively classify the disease with reduced features. Since BABC has a unique solution update mechanism as it allows the results to converge to the optimal solution quickly.

REFERENCES

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

- [2] Alper U., Alper M. and Ratna B.C. , 'mr²PSO:A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification,' *Information Science* 181, pp.4625–4641, 2011.
- [3] Esseghir M.A , '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, 2010.
- [4] Gutlein M, Frank E. and Karwath A, 'Large-scale attribute selection using wrappers,' *In IEEE symposium computational intelligence and data mining*, pp.332–339, 2010.
- [5] Han. J, & Kamber M. Data mining: Concepts and techniques. Morgan Kaufman Publishers, 2001.
- [6] Kahramanli H and Allahverdi N , 'Design of a hybrid system for the diabetes and heart diseases,' *Expert Systems with Applications* 35 pp.82–8, 2008.
- [7] Nahar J.T.I, Kevin S.T and Y.P.chen., 'Computational Intelligence for heart disease diagnosis: A medical knowledge driven approach', *Expert systems with applications* 40, pp.96–104, 2013.
- [8] Lin S.W, Lee Z.J, Chen S.C and Tseng T.Y, 'Parameter determination of support vector machine and feature selection using simulated annealing approach', *Appl. Soft computing* 8, pp.1505–1512, 2008.
- [9] Liu H. and Yu L., 'Toward integrating feature selection algorithms for classification', *IEEE Trans. Knowledge Data Eng*, 2005.
- [10] Pnendharkar P. C, Rodger J. A, Yaverbaum G. J, Herman .N and Benner M, 'Association, statistical mathematical and neural approaches for mining breast cancer patterns', *Expert Systems with Applications* 17, pp.223–232, 1999.
- [11] Song.M.H., Lee.J, Cho.S.P., Lee.K.J and Yoo.S.K, 'Support vector machine Based Arrhythmia classification using reduced Features', *International Journal of control, Automation, and Systems*, Vol 3, pp.571–579, 2005
- [12] Sahu B. and Mishra D, 'A Novel Feature Selection algorithm using Particle Swarm Optimization for cancer Microarray Data', *Procedia Engineering* 38, pp.27–31, 2012
- [13] Subanya B and Rajalaxmi R.R, 'Feature Selection using Artificial Bee Colony for Cardiovascular Disease Classification', *Proc. Of International Conference on Electronics and Communication System, ICECS –2014* , Karpagam College of Engineering Coimbatore , 2014
- [14] Swati S. and Ashok G., 'Feature selection for medical diagnosis: Evaluation for cardiovascular diseases', *Expert Systems with applications* 40, pp.4146–4153 , 2013
- [15] Tang, Y., Jin, B., Sun,Y. & Zhang, Y.-Q, 'Granular support vector machines for medical binary classification problems', *IEEE symposium on computational intelligence in bioinformatics and computational biology* , pp. 73–78, 2004
- [16] Thananan Prasartvit., Anan Banharnsakun., Boonserm Kaewkamnerdpong. and Tiranee Achalakul, 'Reducing bio-informatics data dimension with ABC-KNN', *Neurocomputing* 116, pp. 367–381, 2013
- [17] Vieira.S.M, Luis F. Mendonca.L.F, Farinha.G.J. and Sousa M.C.J, 'Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients', *Applied soft computing* 13, pp.3494–3504, 2013