

An Ensemble Optimization for Heart Disease Classification and Attribute Filtering

K. Sudhakar

Research Scholar,
Department of Computer Science
Shrimati Indira Gandhi College,
Trichy, Tamilnadu, India

Dr. M. Manimekalai

Director,
Department of Computer Science
Shrimati Indira Gandhi College,
Trichy, Tamilnadu, India

Abstract— The more experience must be required for diagnosing the heart disease and it is a complex task. The Heart MRI, ECG and Stress Test etc are the numbers of medical tests are prescribed by the doctor for examining the heart disease and it is the way of tradition in the prediction of heart disease. Today world, the hidden information of the huge amount of health care data is contained by the health care industry. The effective decisions are made by means of this hidden information. For appropriate results, the advanced data mining techniques with the information which is based on the computer are used. Classification is a classic data mining task, with roots in machine learning. The classification techniques like Support Vector Machines (SVM), Naive Bayesian Theorem, Classification tree and Neural Network are used in this paper to know the classification accuracy of the techniques in the prediction of the heart disease. Feature selection methods involve generation of the subset, evaluation of each subset, criteria for stopping the search and validation procedures. The characteristics of the search method used are important with respect to the time efficiency of the feature selection methods. The following attribute filtering techniques CFS, PCA, Information gain. In this paper, we can predict that which classification algorithm can be used to predict the heart disease in earlier stage as well as the attribute filtering techniques which will be suit before the classification to reduce the dataset size and most important attributes can be selected for predicting. From this paper, we can predict that the artificial neural network and PCA filtering method will be best for predicting the heart disease using the dataset.

Keywords—ANN, SVM, CFS, PCA.

I. INTRODUCTION

In the past 10 years Heart disease becomes the major cause for death all around the globe (World Health Organization 2007) [1]. To help the professionals of health care, the several data mining techniques are used by the researchers in the findings of heart disease. The European Public Health Alliance stated that heart attacks, strokes and other circulatory diseases is accounted as 41% of all deaths (European Public Health Alliance 2010) [2]. The one fifth lives of the Asian are lost due to the non communicable disease such as chronic respiratory diseases, cardiovascular diseases and cancer etc and this is described in the ESCAP (Economical and Social Communication of Asia and Pacific 2010) [2]. ABS (The Australian Bureau of Statistics) described that circulatory system diseases and heart diseases are the primary reason for death in Australia, causing 33.7% all deaths (Australian

Bureau of Statistics 2010) [3]. The heart disease patients were motivated around the globe every year. In addition to the availability of large amount of patients' data from which to extort useful knowledge, in the diagnosis of heart disease, for facilitating the professionals of health care, the data mining techniques have been used by the researchers [4]. Nowadays, data mining is the exploration of large datasets to extort hidden and formerly unknown patterns, relationships and knowledge that are complicated to detect with conventional statistical methods. In the emerging field of healthcare data mining plays a major role to extract the details for the deeper understanding of the medical data in the providing of prognosis [5]. Due to the development of modern technology, data mining applications in healthcare consist about the analysis of health care centres for enhancement of health policy-making and prevention of hospital errors, early detection, prevention of diseases and preventable hospital deaths, more value for money and cost savings, and detection of fraudulent insurance claims.

The characteristic selection has been an energetic and productive in the field of research area through pattern recognition, machine learning, statistics and data mining communities. The main intention of attribute selection is to choose a subset of input variables by eradicating features, which are irrelevant or of non prognostic information. Feature selection [6] has proven in both theory and practice to be valuable in ornamental learning efficiency, escalating analytical accuracy and reducing complexity of well-read results. Feature selection in administered learning has a chief objective in finding a feature subset that fabricates higher classification accuracy. The number of feature N increases because the expansions of the domain dimensionality. Among that finding an optimal feature subset is intractable and exertions associated feature selections have been demonstrated to be NP-hard. At this point, it is crucial to depict the traditional feature selection process, which consists of four basic steps, namely, validation of the subset, stopping criterion, evaluation of subset and subset generation. Subset generation is a investigation process that generates the candidate feature subsets for assessment based on a certain search strategy. Depends on the certain assessment, the comparison with the best prior one and each candidate subset is evaluated. If the new subset revolves to be better, it reinstates best one. Whenever the stopping condition is fulfilled until the process is repeated. There are large number

of features that can exceed the number of data themselves often exemplifies the data used in ML [7]. This kind of problem is known as "the curse of dimensionality" generates a challenge for a mixture of ML applications for decision support. This can amplify the risk of taking into account correlated or redundant attributes which can lead to lower classification accuracy. As a result, the process of eliminating irrelevant features is a crucial phase for designing the decision support systems with high accuracy.

In this technical world, data mining is the only consistent source accessible to unravel the intricacy of congregated data. Meanwhile, the two categories of data mining tasks can be generally categorized such as descriptive and predictive. Descriptive mining tasks illustrate the common attributes of the data in the database. Predictive mining tasks execute implication on the present data in order to formulate the predictions. Data available for mining is raw data. The data collects from different source, therefore the format may be different. Moreover, it may consist of noisy data, irrelevant attributes, missing data etc. Discretization – Once the data mining algorithm cannot cope with continuous attributes, discretization [8] needs to be employed. However, this step consists of transforming a continuous attribute into an unconditional attribute, taking only a small number of isolated values. Frequently, Discretization often improves the comprehensibility of the discovered knowledge. Attribute Selection – not all attributes are relevant and so for selecting a subset of attributes relevant for mining, among all original attributes, attribute selection [9] is mandatory.

II. CLASSIFICATION TECHNIQUES

The most commonly used data mining technique is the classification that occupies a set of pre-classified patterns to develop a model that can categorize the population of records at large. The learning and classification is involved by the process called the data classification. By the classification algorithm, the training data are analyzed in the learning [4] [10]. The approximation of the classification rules, the test data are used in the classification. To the new data tuples, the rules can be applied when the accuracy is acceptable. To verify the set of parameters which is needed for the proper discrimination, the pre-classified examples are used in the classifier-training algorithm. The model which is called as a classifier, only after these parameters encoded by the algorithm. The classification tasks are explained in subsequent methods as Artificial neural network, Naive Bayesian classification algorithm classification techniques are deployed in this paper.

A. Support Vector Machine

For the classification and regression, the diagnosis of the medical data uses the set of related supervised learning method is called as the SVM (Support Vector Machine) [11]. SVM can be used for pattern classification and nonlinear regression. Support Vector Machines (SVM's) are a relatively new learning method used for binary classification. The geometric margin maximization and minimizes the empirical classification error is done simultaneously in SVM. SVM is

called Maximum Margin Classifiers and it can be efficiently perform non-linear classification using kernel trick. Using of the RBF (Radial Basis Function) kernel of SVM is like the classifier, the higher dimensional data can be analysed by RBF kernel function. RBF kernel function [12] is used because RBF kernel nonlinearly maps samples into a higher dimensional space and also it has less numerical difficulties. The values fed to the SVM classifier are normalized initially to improve the accuracy. Test data sets were used to assess the performance of the SVM model. Validation using the test data sets avoid potential bias of the performance, estimate due to over-fitting of the model to training data sets. For classification, RBF kernel with the SVM classifier is used. An automated classifier for the discrimination between the person with heart disease and without heart disease has been developed using supervised learning algorithm named SVM. The support vector machine can provide good generalization performance on pattern classification problem. It is considered a good classifier because of its high generalization performance without the need to add a priori knowledge. The aim of SVM is to find the best classification function to distinguish between members of the two classes in the training data.

B. Classification Tree

The decision tree is also known as the classification tree [13]. In this research, the decision tree type used as gain ratio decision tree. The gain ratio decision tree is based on the entropy (information gain) approach, which opts for the dividing the aspects that minimize the value of entropy, thus maximizing the information gain. The information gain is the difference between the original information content and the amount of information required. The features are categorized by the information gains, and then the top ranked features are chosen as the probable attributes used in the classifier. In order to classify the splitting attribute of the decision tree, one must determine the information gain for each attribute and then select the attribute that maximizes the information gain [14]. The information gain for each attribute is calculated using the following formula:

$$E = \sum_{i=1}^k P_i \log_2 P_i$$

Where k is the number of classes of the target attributes P_i is the number of occurrences of class i divided by the total number of instances (i.e. the probability of i occurring). To minimize the effect of bias resulting from the use of information gain, a variant known as gain ratio was introduced by the Australian academic Ross Quinlan. The information gain measure is inclined towards the tests with many outcomes. The attributes with the large number of values is selected and it is preferred in this information gain ratio. Gain Ratio amends the information gain for each attribute to allocate for the breadth and uniformity of the attribute values.

$$\text{Gain ratio} = \text{Information gain} / \text{Split Information}$$

C. Naive Bayesian Classification Algorithm

The Bayesian Classification signifies a supervised learning method and also as a statistical method for classification [15]. In which it assumes a fundamental

probabilistic model and it allocates us to incarcerate uncertainty about the model in a ethical way by determining probabilities of the results. And also it can unravel diagnostic and predictive problems. The Bayesian theorem as follows:

Given training data Y , posterior probability of a hypothesis I is $R(I|Y)$, follows the Bayes theorem $R(I|Y) = R(Y|I)R(I)/R(Y)$.

Algorithm:

The Naive Bayes algorithm is based on Bayesian theorem as given by above equation.

Step 1: Each data sample is represented by an n dimensional feature vector, $S = (s_1, s_2, \dots, s_n)$, depicting n measurements made on the sample from n attributes, respectively S_1, S_2, \dots, S_n .

Step 2: Suppose that there are m classes, T_1, T_2, \dots, T_k . Given an unknown data sample, Y (i.e., having no class label), the classifier will predict that Y belongs to the class having the highest posterior probability, conditioned if and only if:

$R(T_j|Y) > R(T_l|Y)$ for all $1 \leq l \leq k$ and $l \neq j$

$R(T_j|Y)$ is maximized then. $R(T_j|Y)$ is maximized for the class T_j is called the maximum posterior hypothesis. By Bayes theorem,

Step 3: Only $R(Y|T_j)R(T_j)$ need be maximized when $R(Y)$ is constant for all classes. It is assumed that the classes are equally likely when the class prior probabilities are not known, i.e. $R(T_1) = R(T_2) = \dots = R(T_k)$, and consequently we would maximize $R(Y|T_j)$. or else, we maximize $R(Y|T_j)R(T_j)$. Note that the class prior probabilities may be estimated by $R(T_j) = a_j/a$, where A_j is the number of training samples of class T_j , and a is the total number of training samples on Y . That is, the naive probability assigns an unknown sample Y to the class T_j .

D. Artificial Neural Network

Moreover, the realistic provisional terms in neural networks are non-linear statistical data modelling tools. For discovering the patterns or modelling the complex relationships between inputs and outputs the neural network can be used. The process of collecting information from datasets is the data warehousing firms also known as data mining by using neural network tool [15]. The more informed decisions are made by users that helping data of the cross-fertilization and there is distinction between these data warehouse and ordinary databases and there is an authentic manipulation.

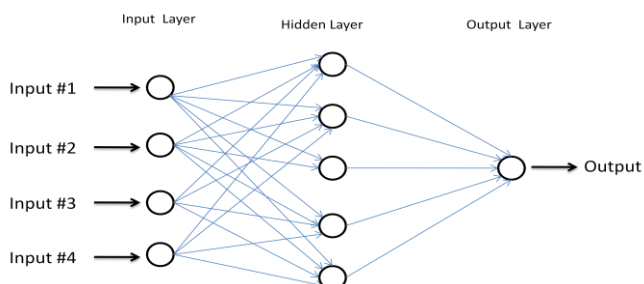


Figure 1: Example Artificial Neural Network

Among the algorithms the most popular neural network algorithms are Hopfield, Multilayer perception, counter propagation networks, radial basis function and self organizing maps etc. In which, the feed forward neural network was the first and simplest type of artificial neural network consists of 3 units input layer, hidden layer and output layer. There are no cycles or loops in this network. A neural network [20] has to be configured to fabricate the required set of outputs. Basically there are three learning conditions for neural network. 1) Supervised Learning, 2) Unsupervised Learning, 3) Reinforcement learning the perception is the basic unit of an artificial neural network used for classification where patterns are linearly separable. The basic model of neuron used in perception is the Mcculluchpitts model. The learning for artificial neural networks are as follows:

Step 1: Let $D = \{(X_i, Y_i) / i=1, 2, 3, \dots, n\}$ be the set of training example.

Step 2: Initialize the weight vector with random value, $W(0)$.

Step 3: Repeat.

Step 4: For each training sample $(X_i, Y_i) \in D$.

Step 5: Compute the predicted output $Y_i^{\wedge}(k)$

Step 6: For each weight we do.

Step 7: Update the weight $w_e(k+1) = W_j(k) + (y_i - y_i^{\wedge}(k))x_{ij}$.

Step 8: End for.

Step 9: End for.

Step 10: Until stopping criteria is met.

III. ATTRIBUTE FILTERING TECHNIQUES

In general, Feature subset selection is a pre-processing step used in machine learning [17]. It is valuable in reducing dimensionality and eradicates irrelevant data therefore it increases the learning accuracy. It refers to the problem of identifying those features that are useful in predicting class. Features can be discrete, continuous or nominal. On the whole, features are described in three types.

1) Relevant 2) Irrelevant 3) Redundant. Feature selection methods wrapper and embedded models. Moreover, Filter model rely on analyzing the general qualities of data and evaluating features and will not involve any learning algorithm, where as wrapper model uses a determined learning algorithm and use learning algorithms performance on the provided features in the evaluation step to identify relevant feature. The Embedded models integrate the feature selection as a part of the model training process.

The collection of data from medical sources is highly voluminous in nature. The various significant factors distress the success of data mining on medical data. If the data is irrelevant, redundant then knowledge discovery during training phase is more difficult. Figure 2 shows flow of FSS.

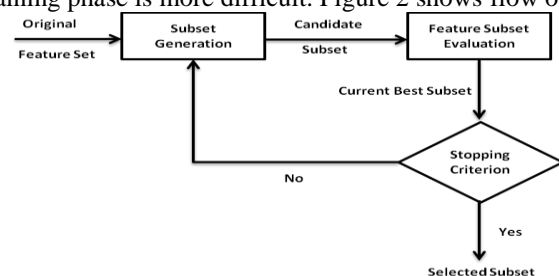


Figure 2: Feature Subset Selection

A. Correlation based Feature Selection Method

Redundancy of attributes in DWs is a main concern. When the attribute are derived from set of attributes or another attributes then that attribute can be redundant and the attribute also be the redundant ones when the attributes are strongly related by some other attributes. Using the correlation analysis, the detection of redundancy takes place [18]. Based on the data, how it is one attributes can be implied by other attribute is determined by such analysis when two attributes are given. By computing the correlation coefficient between two attributes X and Y, then we can access the correlation between them for given numerical attributes.

Finally, as I conclude that when the class concept is relevant to the attribute/feature if it is good and another relevant feature is not redundant. The good feature is developed when it is highly correlated to the class not to any other features by implementing the goodness measure as the association between two variables. The attribute is relevant to the class are made when the correlation among the class and feature/attribute is high and it is does not attain the level when it is relevant to another feature. For the classification task, when the attribute is predicted by the other relevant feature/attributes. The correlation among the attributes and procedure for selecting the attributes which is based on the evaluation and this is necessitated in the problem of attribute selection. This problem is solved by the formula for classical linear correlation between two random variables or attributes. The linear correlation coefficient is given by the following formula:

$$\text{Correlation}(r) = \frac{N \sum XY - \sum X \sum Y}{\sqrt{(N \sum X^2 - (\sum X)^2)(N \sum Y^2 - (\sum Y)^2)}}$$

Where X and Y are the two features/attributes.

B. Principle Component Analysis (PCA) Attribute Filtering Method

The Principle component analysis (PCA) [19] is a stagnant technique used in many applications like face recognition, pattern recognition, image compression and data mining. Further PCA is used to shrink the dimensionality of the data consisting of a large no. of attributes. PCA can be generalized as multiple factor analysis and as correspondence analysis to handle heterogeneous sets of variables and quantitative variables respectively. The following are the main procedure for the principle component analysis (PCA). Scientifically PCA depends on SVD of rectangular matrices and Eigen decomposition of positive semi definite matrices.

Step 1: Obtain the input matrix

Step 2: Subtract the mean from the data set in all dimensions

Step 3: Calculate covariance matrix of this mean subtracted data set.

Step 4: Calculate the Eigen values and Eigen Vector from covariance matrix

Step 5: Form a feature vector

Step 6: Derive the new data set.

C. Information Gain Attribute Filtering Method

In this method, the discernibility function is used [20]. The discernibility function is given as follows: For an information system (I,H), s discernibility function DF is a boolean function of m Boolean variables e_1, e_2, \dots, e_n corresponding to the attributes e_1, e_2, \dots, e_n respectively, and defined as follows: $DF(e_1, e_2, \dots, e_n) = S_1 \wedge S_2 \wedge \dots \wedge S_m$ where $e_j \in S$. The proposed algorithm for the information gain attribute subset evaluation is defined as below:

Step 1: Compute discernibility matrix for the selected dataset.

By using $P[J, I] = \{ e \in E, \text{ where } X[J] \neq X[I] \text{ and } Y[J] \neq Y[I] \}$ $J, I = 1, 2, \dots, m$ Eq1 Where X are conditional attributes and Y is a decision attribute. This discernibility matrix P is symmetric. Where $P[a, b] = P[b, a]$ and $P[a, a] = 0$. Therefore, it is sufficient to consider only the lower triangle or the upper triangle of the matrix.

Step 2: Compute the discernibility function for the discernibility matrix $P[a, b]$ by using $DF(a) = \bigwedge \{ P[a, b] / a, b \in I; P[a, b] \neq 0 \}$ Eq2

Step 3: Select the attribute, which belongs to the large number of conjunctive sets, numbering at least two, and apply the expansion law.

Step 4: Repeat steps 1 to 3 until the expansion law cannot be applied for each component. Step 5: Substitute all strongly equivalent classes for their corresponding attributes.

Step 6: Calculate the Information gain for the simplified discernibility function contained attributes by using $\text{Gain}(G_i) = F(R_i) - F(G_i)$ Eq.3 Where

$$F(G) = \sum_{j=1}^m R_j \log_2 R_j \quad \text{---- (4)}$$

$$\sum_{j=1}^m R_j \log_2 R_j = -\frac{r_1}{r} \log_2 \frac{r_1}{r} - \frac{r_2}{r} \log_2 \frac{r_2}{r} \dots - \frac{r_m}{r} \log_2 \frac{r_m}{r} \quad \text{---- (5)}$$

Where R_j is the ratio of conditional attribute R in dataset. When G_i has $|G_i|$ kinds of attribute values and condition attribute R_j partitions set R using attribute G_i , the value of information $F(G_i)$ is defined as

$$F(G_i) = \sum_{i=1}^{G_i} K_i * F(B_i) \quad \text{---- (6)}$$

Step 7: Choose the highest Gain value and add it to the reduction set, and remove the attribute from the discernibility function. Goto step 6 until the discernibility function reaches null set.

IV. DATASET AND IMPLEMENTATION TOOL

The data used in this study is the Hungarian institute of cardiology. The dataset contains 303 instances and 14 attributes of the heart disease patient [21]. The general purpose machine learning and data mining tool is an Orange (<http://orange.biolab.si>). It features a multilayer architecture suitable for different kinds of users, from inexperienced data mining beginners to programmers who prefer to access the tool through its scripting interface. In the paper we outline the history of Orange's development and present its current state, achievements and the future challenges. The following are the dataset attributes.

S.No	Attribute
1	Age
2	Gender
3	Chest Pain
4	Rest SBP
5	Cholestrol
6	Fasting Blood
7	Rest ECG
8	Maximum HR
9	Exer Ind
10	ST by exercise
11	Slope peak exc ST
12	Major vessels colored
13	Thal
14	Diameter Narrowing

TABLE 1: THE HEART DISEASE DATASET

V. EXPERIMENTAL RESULT

In this experiment, we can examine the Classification accuracy (CA), Sensitivity (Sens), Specificity, Area under ROC Curve, Information Score (IS), F-Measure (F1), Precision (Prec), Recall, Brier Score and Matthews Correlation Coefficient.

Methods	CA	Sens	Spec	AUC	IS	F1	Prec	Recall	Brier	MCC
Classification Tree	0.7523	0.7866	0.7122	0.7960	0.4917	0.7748	0.7633	0.7866	0.4270	0.5005
SVM	0.8383	0.8537	0.8201	0.9051	0.5277	0.8511	0.8485	0.8537	0.2463	0.6742
Naive Bayes	0.8285	0.8476	0.8058	0.8956	0.6085	0.8424	0.8373	0.8476	0.2770	0.6541
Neural Network	0.8123	0.7564	0.6922	0.7850	0.4752	0.7580	0.7522	0.7788	0.2398	0.4836

TABLE 2: THE EXPERIMENTAL RESULT OF HEART DISEASE DATASET FOR PREDICTED CLASS 0

Methods	CA	Sens	Spec	AUC	IS	F1	Prec	Recall	Brier	MCC
Classification Tree	0.7523	0.7122	0.7866	0.7960	0.4917	0.7253	0.7388	0.7122	0.4270	0.5005
SVM	0.8383	0.8201	0.8537	0.9051	0.5277	0.8231	0.8261	0.8201	0.2463	0.6742
Naive Bayes	0.8285	0.8058	0.8476	0.8956	0.6085	0.8116	0.8175	0.8058	0.2770	0.6541
Neural Network	0.7453	0.7058	0.7780	0.7839	0.4845	0.7178	0.7262	0.7014	0.2429	0.4740

TABLE 3: THE EXPERIMENTAL RESULT OF HEART DISEASE DATASET FOR PREDICTED CLASS 1

	0	1	Instances
0	78.7%	21.3%	164
1	28.8%	71.2%	139
Instances	169	134	303

TABLE 4: CONFUSION MATRIX FOR CLASSIFICATION TREE

	0	1	Instances
0	85.4%	14.6%	164
1	20.9%	79.1%	139
Instances	165	138	303

TABLE 5: CONFUSION MATRIX FOR SVM

	0	1	Instances
0	84.8%	15.2%	164
1	19.4%	80.6%	139
Instances	166	137	303

TABLE 6: CONFUSION MATRIX FOR NAIVE BAYES CLASSIFICATION

	0	1	Instances
0	87.8%	12.2%	164
1	18.0%	82.0%	139
Instances	173	130	303

TABLE 7: CONFUSION MATRIX FOR NEURAL NETWORK

- 1) Selected Attribute using Information gain: 3 (Chest pain), 7 (Rest ECG), 11 (Slope peak ex ST), 13 (Thal), 2 (Gender), 6 (Fasting Blood sugar), 9 (exerc ind ang)
- 2) Selected Attribute using CFS: 3 (Chest pain), 7 (Rest ECG), 11 (Slope peak ex ST), 13 (Thal), 2 (Gender), 6 (Fasting Blood sugar), 9 (exerc ind ang)
- 3) Selected Attribute using PCA: 1 (Age), 3 (Chest pain), 4 (Rest SBP), 5 (Cholestrol), 7 (Rest ECG), 8 (Max HR), 10 (ST by exercise), 12 (Major Vessels Colored).

Feature Selection Method	Number of Selected Attributes	Total number of attributes
Information Gain	7	14
Correlation based Feature Selection	6	14
PCA	8	14

TABLE 8: NUMBER OF SELECTED ATTRIBUTES FROM THE GIVEN NUMBER OF ATTRIBUTES USING ATTRIBUTE FILTERING METHOD

VI. ANALYSIS OF THE RESULT

From the above result of table 2 and 3, we can conclude that the classification accuracy (AC), Sensitivity, Specificity, Area under ROC curve, Information score, F-Measure, Precision, Recall, Brier score and Matthews Correlation Coefficient (MCC) are less in the Artificial neural network (ANN) than SVM, classification tree and Naive Bayes classification method. The table 4,5,6 and 7 shows that the confusion matrix of proportions of true percentage for ANN, SVM, Naive Bayes classification and classification tree. The correctly classified percentage of neural network is greater than the others. The correctly classified instances percentage are 87.8% for class 0 and 82.0% for class 1, whereas the percentage of misclassified instances also least for neural network when it is compared with others. The percentage for misclassified instances of neural network is 12.2% and 18.0%. From the above results we can conclude that the neural network will be suitable for the classification task for given heart disease dataset. From the above result of feature selection, 7 attributes are filtered using information gain attribute filtering method, 8 attributes are selected using PCA and only 6 attributes are filtered in CFS out of 14 attributes. From this result, we can conclude that, Correlation based Feature Selected (CFS) method is best to predict the heart disease for the given data set than PCA and Information gain attribute filtering.

VII. CONCLUSION

In this paper, we have done a comparative study of the classification techniques and attribute filtering techniques (Feature selection techniques) for predicting the heart disease by using given dataset. The model for predicting in the data mining is always referred as the classification techniques. In this paper, the classification tree, Support Vector Machine (SVM), Artificial Neural Networks and Naive Bayesian classification are examined and the experimental shows that the artificial neural network is best to predict the heart disease in terms of classification accuracy, sensitivity, specificity, area under ROC curve, precision, recall, brier score, information score, MCC. Feature selection technique is used to reduce the execution by selecting the important attributes in terms of some conditions. The attribute filtering methods like PCA (Personal Component Analysis), CFS (Correlation based Feature Selection) and information gain attribute filtering. From these methods, CFS produces least number of instances than the PCA and information gain attribute filtering.

REFERENCES

- [1] A Safe Future- Global Public Health Security in the 21st Century, The World Health Report 2007.
- [2] Commission Staff Working Document, European Commission, Implementation of Health Programme 2010.
- [3] Drugs in Australia 2010-Tobacco, Alcohol and other drugs, Australian Institute of Health and Welfare.
- [4] Vikas Chaurasia, Saurabh Pal, Early Prediction of Heart Diseases Using Data Mining Techniques, Carib.j.SciTech,2013,Vol.1, pp.no:208-217.
- [5] Hian Chye Koh and Gerald Tan, Data Mining Applications in Healthcare, Journal of Healthcare Information Management — Vol. 19, No. 2, pp.no: 64-72.
- [6] S.Saravanakumar, S.Rinesh, Effective Heart Disease Prediction using Frequent Feature Selection Method, International Journal of Innovative Research in Computer and Communication Engineering, Vol.2, Special Issue 1, March 2014, pp.no: 2767-2774.
- [7] Jyoti Soni, Uzma Ansari, Dipesh Sharma and Sunita Soni, Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers, International Journal on Computer Science and Engineering (IJCSE), Vol. 3 No. 6 June 2011, pp.no:2385-2392.
- [8] Aieman Quadir Siddique, Md. Saddam Hossain, Predicting Heart-disease from Medical Data by Applying Naïve Bayes and Apriori Algorithm, International Journal of Scientific & Engineering Research, Volume 4, Issue 10, October-2013, pp.no: 224-231.
- [9] Priyanka Palod, Jayesh Gangrade, Efficient Model For Chd Using Association Rule With Fds, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 2, Issue 6, June 2013, pp.no:2157-2162.
- [10] Shamsheer Bahadur Patel, Pramod Kumar Yadav, Dr. D. P.Shukla, Predict the Diagnosis of Heart Disease Patients Using Classification Mining Techniques, IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS), Volume 4, Issue 2 (Jul. - Aug. 2013), PP 61-64.
- [11] Ashfaq Ahmed K, Sultan Aljahdali and Syed Naimatullah Hussain, Comparative Prediction Performance with Support Vector Machine and Random Forest Classification Techniques, International Journal of Computer Applications (0975 – 8887) Volume 69– No.11, May 2013, pp.no:12-16.
- [12] R. Chitra and Dr.V. Seenivasagam, Heart Disease Prediction System Using Supervised Learning Classifier, Bonfring International Journal of Software Engineering and Soft Computing, Vol. 3, No. 1, March 2013, pp.no: 1-7.
- [13] ShravanKumar Uppin, Anusuya M A, Expert System Design to Predict Heart and Diabetes Diseases, International Journal of Scientific Engineering and Volume No.3 Issue No.8, 1 August 2014, pp : 1054-1059.
- [14] Aswathy Wilson, Jismi Simon, Liya Thomas, Soniya Joseph, Data Mining Techniques For Heart Disease Prediction, International Journal of Advances in Computer Science and Technology, Volume 3, No.2, February 2014, pp.no: 113-116.
- [15] D Ratnam, P HimaBindu, V.Mallik Sai, S.P.Rama Devi, P.Raghavendra Rao, Computer-Based Clinical Decision Support System for Prediction of Heart Diseases Using Naïve Bayes Algorithm, International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, pp.no: 2384-2388.
- [16] Ashish Kumar Sen, Shamsheer Bahadur Patel, Dr. D. P. Shukla, A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach Two Level, International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 2 Issue 9 Sept., 2013 pp.no: 2663-2671.
- [17] M. Anbarasi, E. Anupriya, N.CH.S.N.Iyengar, Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm, International Journal of Engineering Science and Technology, Vol. 2(10), 2010, 5370-5376.
- [18] Dr.B.Sarojini, A Wrapper Based Feature Subset Evaluation Using Fuzzy Rough K-NN, International Journal of Engineering and Technology (IJET), Vol 5 No 6 Dec 2013-Jan 2014, pp.no:4672-4676.
- [19] Negar Ziasabounchi and Iman N. Askerzade, A Comparative Study of Heart Disease Prediction Based on Principal Component Analysis and Clustering Methods, Turkish Journal of Mathematics and Computer Science, pp.no:1-11.
- [20] Selvakumar.P, DR.Rajagopalan.S.P, A Survey On Neural Network Models For Heart Disease Prediction, Journal of Theoretical and Applied Information Technology, 20th September 2014. Vol. 67 No.2, pp.no:485-497.