Predication of Arrhythmia Disease using Data Mining Classification Algorithm with Feature Selection Techniques

Sunil Kumar Saini[1] Mr. Dinesh Swami[2] Dr. Saroj Hiranwal[3]
[2]Assistant Professor [3]Principal
[1,2,3]Department of Computer Science Engineering
[1,2,3]RIET College Jaipur, India

Abstract— This Paper aimed to find Which classification algorithm has the highest accuracy for the medical dataset arrhythmia before feature selection, which feature selection algorithm among filter and wrapper gives the best result, is feature selection technique of preprocessing improves the performance accuracy of arrhythmia dataset and how accurate is mining technique in predicting the person suffering from arrhythmia disease.

Keywords: Data Mining, Arrhythmia Disease

I. INTRODUCTION

This chapter gives introduction to data mining and classification techniques, used for the purpose of analyzing the sample dataset. It throws light on scope, objective and research hypothesis for the arrhythmia disease data.

Data mining is a term which refers to extracting or "mining" knowledge patterns from huge quantity of data. "Knowledge mining," a shorter term, may not reflect the emphasis on mining from large amounts of data. Mining characterizes the process in finding a small set of precious nuggets from a great deal of raw material. Basically, data mining attempts to extract hidden patterns and trends from large databases and also supports automatic exploration of data. From these patters some rules are derived, which enable the users to examine and review their decisions related to business or scientific area, thus resulting in more interaction with databases and data warehouses.

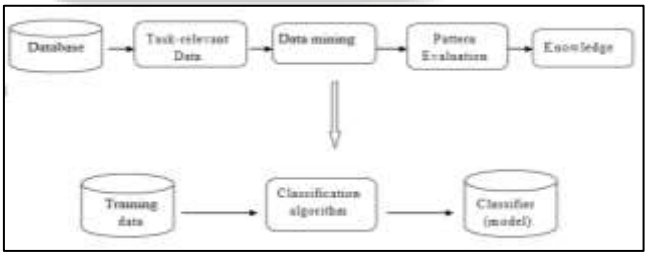


Fig. 1.1: Knowledge Discovery

Data mining is a core of knowledge discovery process shown in fig 1.1. As data are drawn from multiple sources, they may contain inconsistent, incorrect or missing information. For example, it is very much possible that the same information in different sources can be presented in different formats. Data transformation process converts or consolidates the data.

Into suitable format for the data mining process by performing aggregation operations. Task-relevant data is needed to be prepared before going to data mining process. Data mining phase is concerned with the process of extracting patterns from a huge amount of data. Based on the data mining task being performed, this step applies algorithms to the transformed data to generate the desired output. Pattern evaluation identifies the interesting patterns obtained in the data mining stage and converting them into knowledge. This knowledge is then used by users in decision-making. Knowledge discovery makes use of data visualization and knowledge representation techniques to present the data mining results to users. The process of data mining consists of three major stages [5]-

- Exploration
- Model building and validation
- Deployment

The main goal of data mining is to make the knowledge, extracted from the data into human understandable structure. Exploration is the initial phase of mining to prepare the data or data is cleaned and transformed, to make it available for further processing. The next phase picks out the model from all which gives the best predictive performance as there is variety of techniques available to reach the same goal. Finally, the model is deployed in last stage to make approximations for the expected outcome.

Classification is a data mining task of assigning the object to predefined categories. It is done by building a model which is based on one or more categorical variables or attributes. The process begins with the training data attribute set and maps it into predefined class label. The modeling is done for two purposes. Descriptive modeling is a process describing the real world events and relationship between them. Predictive modeling is used to predict the outcome or unknown events. Classification applications usually have more than one class of data. Two types of data are considered here. Training data is a set of sample points that we have as of now. Holdout set is the subset removed from training set, for reducing the complexity of data. Test data is a set whose class label is to be predicted. It is used to access generalization error of final model described.

The ECG or electro-cardiogram is the signal which measures electrical activity of the heart cells. Arrhythmia is a heart disease in which heart rate becomes irregular. Various patterns are generated using various readings of these waves. To classify it, various researchers had used classifying techniques. Classification techniques are used to analyze the data to categorize them in classes.

II. DECISION TREE

Decision tree induction is a technique used for mining large database addressing efficiency and scalability issues composed of generalization by attribute oriented induction, relevance analysis and multilevel mining [1]. It is accepted as the best technique because it is easy to understand as well as has clarification and flexibility [2]. Techniques to build decision tree include Iterative Dichotomiser tree, C4.5 and CART. Every technique has its own capabilities to deal with different parameters of data as C4.5 is more beneficial than ID3 [2]. It is a supervised learning approach [3].A machine researcher named J. Ross Quinlan in 1980 developed a

decision tree algorithm known as ID3 (Iterative Dichotomiser).Decision tree builds classification model in the form of tree like structure. The leaf node is assigned class label and internal nodes consist of conditions which separates the data having different characteristics. Root node is the top most decision node in a tree.

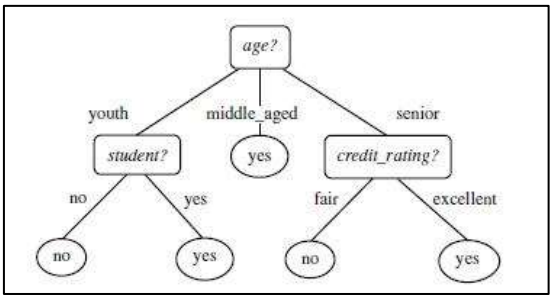


Fig. 1.2: A decision tree for the concept buys computer, indicating whether a customer at All Electronics is likely to purchase a computer. Each internal (non-leaf) node represents a test on an attribute. Each leaf node represents a class (either buys computer = yes or buys computer = no).

ID3 is a decision tree algorithm. It employs top-down approach, greedy search with no backtracking. For construction, it uses entropy and information gain. If entropy is zero, the data sample is completely homogeneous and if entropy is 1, it is equally divided.

1) Calculate entropy of the target.
2) The dataset is then split on the different attributes. The entropy for each branch is calculated. Then it is added proportionally, to get total entropy for the split. The resulting entropy is subtracted from the entropy before the split. The result is the Information Gain, or decrease in entropy.

$$Gain(T,X)= Entropy(T)- Entropy(T,X)$$
$$Entropy(T)=-\Sigma p(x)\log_2 p(x)$$

3) Choose attribute with the largest information gain as the decision node.
4) A branch with entropy of 0 is a leaf node. A branch with entropy more than 0 needs further splitting.
5) The ID3 algorithm is run recursively on the non-leaf branches, until all data is classified.

The cost complexity is measured by the following two parameters number of leaves in tree and error rate of tree. Devi Prasad Bhukya *ET. AL* proposed technique of classification using an AVL tree. It will eventually enhance the quality and also the stability. It has shown how it generalizes the particular hierarchy concept from the training data which is raw using attribute-oriented induction (AOI).Compared various decision making algorithm like CART, ID3, C4.5, SLIQ and SPRINT [1].

A. Naïve Bayesian Classifiers

Bayesian are statistical classifiers based on Baye's theorem. It is used for both supervised as well as unsupervised data [4]. Predicts the probability of the data item belongs to particular class. Prior means all the information from day to day past experience likelihood is the possibility information. Posterior is predicting the particular information from the given set. Evidence is the total number of cases when an event occurs alone.

P(H|X) is the posterior probability of H conditioned on X and P(H) is prior probability.

$$P(H|X)=P(X|H)P(H)/P(X)$$

Here, x_k is value of attribute A_k for tuple X. If A_k is categorical, $P(X|H)= P(x_k|H_i)/ |H_i,D|$ where D denotes the database. If A_k is continuous valued, Gaussian distribution is assumed so that $P(x_k|H_i)=g(x_k ,\mu_{Hi} ,\sigma_{Hi})$.

Bayesians have minimum error rate as compared to other classifiers. It provides theoretical justification as it explicitly uses Bayes theorem. But on the other side it owns to inaccuracy when assumptions are made to reduce the complexity as in class-conditional independence.

Laplacian corrections are used to avoid the conditions when probability value is zero.

The naïve Bayesian classifier makes the assumption of class conditional independence, that is, given the class label of a tuple, the values of the attributes are assumed to be conditionally independent of one another. Bayesian belief networks specify joint conditional probability distributions. They allow class conditional independencies to be defined between subsets of variables. A belief network is defined by two components—a directed acyclic graph and a set of conditional probability tables. Each node in the directed acyclic graph represents a random variable. The variables may be discrete or continuous-valued. Each variable is conditionally independent of its non-descendants in the graph, given its parents. The CPT for a variable Y specifies the conditional distribution P(Y| Parents(Y)), where Parents(Y) are the parents of Y.

B. Rule based Classification

A rule based represents bits of knowledge. If-then rule are used for classification. The condition may consist of more than one attribute sets. Coverage and accuracy accesses rule R on data D.

IF condition THEN conclusion.

An example is rule R1,

R1: IF age = youth AND student = yes THEN buys computer = yes.

The "IF"-part (or left-hand side)of a rule is known as the rule antecedent or precondition. The "THEN"-part (or right-hand side) is the rule consequent.

Data can be uncertain or incomplete.[6] Conflict resolution strategy is needed to figure out which rule is to be assigned for class prediction if more rules are triggered. Priority is assigned having toughest requirement with maximum attribute test under size ordering. In class-based ordering, order of prevalence is considered.

Rule based classifier has the characteristics of mutual exclusion (one rule one record) and exhaustiveness (at least one rule one record). Direct classification rule method extract rules directly from data such as RIPPER. In this method, a rule is learned each time and data covering it are removed. Whereas in indirect methods data is extracted from other classification models such as decision trees or C4.5 rules. C4.5 follows class-based ordering scheme, grouping rule set for a single class and then ranking of rule sets are determined. It orders the rule set to minimize false-positive errors (predicting the class which is not actual in C). These rules select the training data sets not covered in any rule.

A rule R can be assessed by its coverage and accuracy. Given a tuple, X, from a class labeled data set, D, let ncovers be the number of tuples covered by R; ncorrect be the number of tuples correctly classified by R; and |D| be the number of tuples in D. We can define the coverage and accuracy of R as

coverage(R) =ncovers / |D|

accuracy(R) =ncorrect / ncovers

While extracting rules for some data, some inconsistent rules might occur due to the presence of data noise, shortage of information or improper selection of split-points during discretization. Discrete split-point reselection and reducing inconsistent rules by adding attribute [paper] are used for last two problems.

Decision tree and rule-based classification are listed in examples of eager learners as it maps the input attributes to class label as soon the data is available.

C. Nearest neighbor classifier

Classification framework involves two processes, firstly constructing classification model from data and then using that model to test examples. Nearest neighbor classifier is one such approach. Here k-nearest neighbor refers to k points that are close to a data point. This algorithm is known by example-based reasoning, case-based reasoning, memory-based reasoning, instance-based learning and lazy learning [ppt]. It computes the distance or similarity between the test and the training example with in a data. The value of k matters a lot in computing data. Small value of k leads to over fitting and large value cannot classify the data properly. Indexing techniques are used to reduce the complexity while computing the data. The algorithm is basically used for the attributes that are continuous.

D. Artificial Neural Network

Neural network learning algorithm is back-propagation algorithm. Like human brains, ANN consists of nodes and directed links.

PERCEPTRON: The perceptron consists of input nodes and the output nodes to represent the model structure. Output value (y) is calculated the finding the sum of input nodes (x_i) and subtracting the bias value (t) from it. Value of bias depends on sign of output.

$$Y=\text{sign}\ [w_i x_i + w_{i-1}x_{i-1} + \ldots.. + w_1 x_1 - t]$$

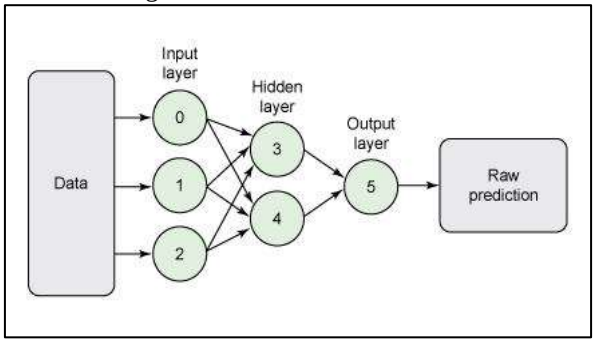


Fig. 1.3: A multilayer feed-forward neural network.

In Multi-layer perceptron, more than one hidden layers are present. Non-linear models can be learned using it.

$$w \leftarrow w - \eta(\alpha \frac{\partial R(w)}{\partial w} + \frac{\partial Loss}{\partial w})$$

Stochastic Gradient Descent (SGD) is used to update the parameters according to the adaptation. Here η denotes learning rate, it controls the step size of parameter search apace. Finding the error rate of hidden nodes is not trivial, for this back-propagation is used. It can be done in forward (weights obtained from previous iteration) as well as backward phase (weight formula is applied in reverse direction).

The characteristics of the data set are examined by identifying the factor that is associated with the results obtained. Statistical research has become a complement in field of medical. In this report, we are identifying the factors on which the arrhythmia data is more correlated using PCA and factor analysis. Decision trees also help in analyzing such factors. Eigen values are calculated and predicting the behavior of chi-square correlation and p-values. To make the model easier to interpret and reduce the over fitting, feature selection is another preprocessing step selecting subset of features including filter and wrapper methods.

III. LITERATURE REVIEW

Data available in many fields is constantly generating, handle such data is very difficult like stock market. Classification techniques are used to solve such problems. Firstly, the information which is required is extracted and then classified [15].

Devi Prasad Bhukya et. al proposed technique of classification using an AVL tree. It will eventually enhances the quality and also the stability. It has shown how it generalizes the particular hierarchy concept from the training data which is raw using attribute-oriented induction (AOI).Compared various decision making algorithm like CART, ID3, C4.5, SLIQ and SPRINT[1]. A machine researcher named J. Ross Quinlan in 1980 developed a decision tree algorithm known as ID3 (Iterative Dichotomiser).Decision tree builds classification model in the form of tree like structure. The leaf node is assigned class label and internal nodes consist of conditions which separates the data having different characteristics. Root node is the top most decision node in a tree. ID3 is a decision tree algorithm. It employs top-down approach, greedy search with no backtracking. For construction, it uses entropy and information gain. If entropy is zero, the data sample is completely homogeneous and if entropy is 1, it is equally divided. Improved ID3 is based on importance of attribute and weight [3].

Kasra Madadipouya has proposed a new algorithm for mining the data which belongs to medical field based on C4.5. It improves the accuracy of classification in comparison to C4.5.During splitting, the proposed algorithm selects the two attributes, using which greater information gain is discovered[2].

David Heckerman has proposed Bayesian networks for data mining. It is used for both supervised as well as unsupervised data[4]. Bayesian are statistical classifiers based on Baye's theorem. Predicts the probability of the data item belongs to particular class. Prior means all the information from day to day past experience likelihood is the possibility information. Posterior is predicting the particular information from the given set. Evidence is the total number

of cases when an event occurs alone. P(H|X) is the posterior probability of H conditioned on X and P(H) is prior probability.

A rule based represents bits of knowledge. If-then rule are used for classification. The condition may consist of more than one attribute sets. Coverage and accuracy accesses rule R on data D. Data can be uncertain or incomplete [6]. Conflict resolution strategy is needed to figure out which rule is to be assigned for class prediction if more rules are triggered. Priority is assigned having toughest requirement with maximum attribute test under size ordering. In class-based ordering, order of prevalence is considered. Rule based classifier has the characteristics of mutual exclusion (one rule one record) and exhaustiveness (at least one rule one record).

LAN Yidong has proposed two methods discrete split-point re-selection, reducing inconsistent rules by adding attribute for removing inconsistency of decision tables. Improvement in mean accuracy is shown in experiment results[5]. M.Akhil jabbaret al. has proposed a new algorithm by combining K-nearest neighbor with genetics(heart disease) getting better classification results[7]. Classification framework involves two processes, firstly constructing classification model from data and then using that model to test examples. Nearest neighbor classifier is one such approach. Here k-nearest neighbor refers to k points that are close to a data point. DAWEN XIA et al. had proposed a new K-NN approach for large datasets to solve the problems that occur in real time predictions [8].

J. S. Raikwal et al. had used SVM and K-NN for performance evaluation and concluded that K-NN is best fitted for numerical data. In other case, it performs with poor results[9]. Like human brains, ANN(artificial neural network) consists of nodes and directed links. The perceptron consists of input nodes and the output nodes to represent the model structure. Output value (y) is calculated the finding the sum of input nodes (x_i) and subtracting the bias value (t) from it. Value of bias depends on sign of output. Neural networks has the ability to handle large data efficiently including reporting, detecting or assimilating[10]. It provides high accuracy, noise tolerance, independence from prior assumptions [11].Mining step consists of two steps data preparation and then rule extraction which is based on feed-forward network. Removes redundant data, discretizes the value of output and produces pruned rules[12]. Support Vector Machines(SVM) are used for describing the decisions taken at boundaries which are based on decision planes. Sets having different membership are partitioned by decision planes. It can also manage multiple continuous and categorical variables [13]. Parallel selective sampling(PSS) is a SVM which is used to reduce imbalance in large data sets, for that data is selected from majority class[14].

Now, the survey is focused on applying the classification techniques on different medical dataset. Mining has played a very significant role in this field technically identifying the patterns which is useful in making the decisions beneficial for clinical purpose.

Indu Saini et al. has used multilayer feed-forward network to achieve good accuracy in classifying various classes of arrhythmia. Performance is evaluated using confusion matrix and other statistical parameters [1]. Alaa Elsayyad et al. imputed missing values in data using expectation maximization and improved feature selection with Pearson Chisquare test. C5.0 boosted decision tree outperformed in diagnosis and can be efficiently used in real world analysis [2]. Feature selection being an active research area nowadays.

Narendra Kohli et al. has found much better accuracy results using feature selection with principal component analysis[3]. Kavita Choudhary evaluated RCT which can help before consulting doctor using decision tree with cross validation, defining most probable factor age with 93.06% accuracy [4].

Abdelghani Bellaachia presented the rate of survive of breast cancer patients. The preliminary results are more promising for data mining applications, predicting problem solutions in medical [5]. Data mining has uncovered a lot of knowledge related to biomedical and healthcare which helped in clinical decision making. As if now, health related data are not confined to quantitative data, so there is a need to explore mining in healthcare.

Parvathi et al. has proposed a hybrid methodology including mining techniques and association rules in enhancing the accuracy [6]. Each algorithm shows different behavior which results that it is infeasible to apply same algorithm for different data. Feature extraction provides the accuracy of classification applied on dataset of gene extraction [7].

A.S.Aneeshkumar et al. has stated that sequential cover approach is effective in health care also apart from other fields with reverse technique to generate appropriate rules for conducting study and analysis on fatty liver disorder [8]. S.Sasikala et al. proposed Shapely Value Embedded Genetic Algorithm (SVEGA) improving breast cancer diagnosis accuracy. It ranks the genes on the basis of its differentiability capability [9].

Ilayaraja M predicted the risk level of people suffering from heart disease. The devised method includes frequent itemsets based on selected symptoms and support values [10].

G.Nagarajan et al. has proposed a hybrid image retrieval system of medical data of cancer, tumour and thyroid [11]. The processing includes three phases- firstly extracting the feature based on gradient extraction algorithm, secondly hybrid approach is used combining branch and bound and artificial bee colony algorithm and lastly feedback method is used improving the performance of hybrid method.

Tapas Ranjan Baitharu proposed six algorithms in classifying the liver disorder and concludes comparing the factors of effectiveness and rate of correction between them [12]. Shubpreet Kaur et al. focused on future trends of KDD using various mining tools for health related issues [13]. The research has found drugs as the most important reason behind having diseases. It also move around analysis of better health related policies and preventing deaths.

Fawzi Elias Bekri et al.investigated KDD in healthcare. It proceeds by saying that onset of disease should be identified in a non-invasive manner [15]. Parvez Ahmad et al. random sampling overcomes all the disadvantages that are faced in clustering process including partitioning as well as hierarchical. It can be due to not knowing the number of clusters, high computational time or large data set [14].

Marios Anthimopoulos et al. has proposed deep convolutional neural network (CNN) for interstitial long disease(ILD) having 5 layers. It captures low level information of features of lung disease. Minimized cross entropy performed for training with Adam optimizer [16]. Riccardo Bellazzi has proposed a framework to handle the various problems which occur while exploring mining models like constructing or assessing the clinical medicine. Discovers non-trivial relationships , rule based models gives proper prediction with bridge model [17]

Monica Adya has raised two issues which can improve the significance of research in this arena. Strong Quality of data is needed to be created so that it can help in conducting research with performance. Partnership between the organizations that are maintaining the data for research and the institutions that are using it should be positive with privacy requirements.

After survey, has came to the conclusion that a hybrid or a combination of more than one data mining techniques can yield much better results than a single technique in health sector. IVF treatment give 70% success rate [19].

Mohammed Abdul Khaleel et al. has used data mining in discovering local frequent diseases by finding pattern in terms of various parameters like accuracy, cost, speed and performance. Mining using medical data also needs business intelligence to support the diagnosis and decisions made on data [20].

IV. PROBLEM STATEMENT

By considering the literature survey of arrhythmia dataset, there is a requirement of feature selection to improve the accuracy. Hybrid approach is also beneficial as it gives appropriate feedback. Other than cost, feature selection is another parameter which is needed to be included in enhancing the performance analysis. To support the findings, predictive analysis using decision tree has improved the performance significantly. Statistical approach using linear discriminate and optimal regression analysis is to generate the rules of diagnosis with appropriate accuracy. In filter feature selection method, PCA is one such statistical approach which has not been used for arrhythmia dataset to that extent. It has proven very good results and suggests that it should be incorporated in final design to enhance the performance of arrhythmia dataset.

V. METHODOLOGY

Various methods used for performance analysis of arrhythmia data are discussed. It gives a view about the feature selection techniques and decision tree algorithm used for the purpose.

A. Decision Tree

Classification has been considered as the most important block for mining the data. It finds the common properties from a set of objects, classifies them and useful patterns are discovered [21]. Decision tree is one of the support tools for making decisions and provides a strategy to reach the goal. Each internal node denotes an attribute on which testing is performed and branch is the result of that test. It implicitly performs feature determination processing. Dynamic Pruning is used for balancing the height using AVL trees depending on priority checks for each node which uses the concept of node merge [21]. For the preparation of data, there is no extra effort needed for the processing. For reducing the depth of trees and accuracy, one or two attributes are chosen as splitting criteria for yielding large information gain ratio. It also enhances probability of finding optimal solution globally [22]. Decision trees are beneficial in handling variety of data as nominal, numeric or textual. It internally processes the data having errors and missing values [23]. The trend of mining with decision trees in healthcare has increased as this sector is full of data and information. The patterns evaluated by practitioners for forecasting, diagnosing and in treating the patients in healthcare

It provides customer oriented approach from which knowledge is being generated. HDSS (Healthcare Decision Support System)is a computer software designed for assisting the physicians at the point of care [25]. Analysis on decisions made is necessary when the actions that has taken lead to conflicting consequences [26].

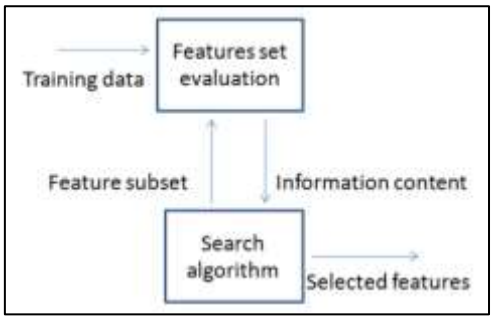
B. Data Preprocessing

The data available in today's database is highly inclined toward various problems like missing value, inconsistent data and noise due to its huge size and variable sources. To improve the quality of mining results, need to improve the quality of data obtained. There are many techniques that can be applied to get rid of these problems. These are as follows:

1) Data cleaning-It is applied on the data to remove the noise and inconsistencies present in the data.
2) Data integration- As the data is collected from variable sources, need to merge and keep it in single store called the data warehouse.
3) Data transformation- Normalize the data as per the requirement of analysis removing the redundancy present and enhancing the accuracy of performance analysis.
4) Data reduction- This can be applied to clearly reduce the data by applying aggregation, eliminating the features which show redundant behavior

C. Feature Selection

Preprocessing is an important process before applying any mining technique. Feature selection is a process of selecting subset of features. It is different from feature extraction in which new features are created from the functions of original features.



For feature subset selection, firstly training data is applied to the evaluation process. Variables are evaluated on the basis of variance, correlations, standard deviation, eigen values or any other statistical means. Then, the information

content is transferred to the algorithm to be applied on values obtained as shown in fig.3.1. This cycle is continued until the proper subset of variables is selected from large dataset. Now, the selected features are ready for further processing of analysis of data.

Importance of feature selection

- Simplification of model to make them easier to interpret.
- Shorter training time
- Avoid the curse of dimensionality.
- Enhanced generalization by reducing over fitting.
- Large irrelevant feature increases computational complexity.
- Gives good accuracy results.

Feature selection is very effective role in dimensional reduction in retaining the information content of data and hence attains the following advantages [78]:

- Better model understanding and visualization- Visualizing with less number of features will result in more clear understanding of the model.
- Generalization of model- Over fitting reduction leads to more accurate learning.
- Efficiency in terms of time and space complexity- Both gets reduced while execution.

VI. FILTER METHODS

Filter methods are generally used as a preprocessing step. The selection of features is independent of any machine learning algorithms. Instead, features are selected on the basis of their scores in various statistical tests for their correlation with the outcome variable. The correlation is a subjective term here. It includes LDA, ANOVA, chi-square, PCA.

To use filter-based feature selection, a target attribute is chosen. It uses statistical measures for predicting the highest power of subset. A score values are calculated within a module to get the desired or relevant attributes. The significant subset of attribute is selected having the best relevance.

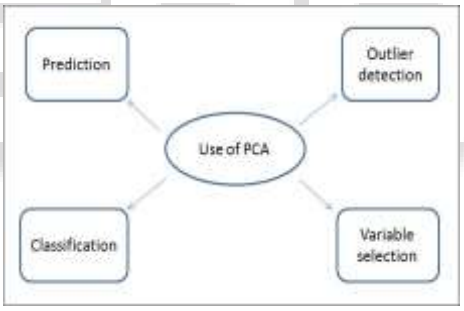
A. PCA

Principal Component analysis is used for transforming possibly correlated features to linearly correlated variables. Its components are always less than the original variables. First variance has largest variance value and so on. It is used for the purpose of dimensional reduction. Interpretation is another goal by using PCA, discovering features from large data set. Eigen-values denote a number which shows how much variance is towards a particular direction. It is a technique used for pre-processing and visualizing the data. It identifies two dimensional planes that describe the highest variance of data optimally [33]. With in it, the original data space gets rotated towards the highest variance direction. PCA in conjunction with regression gives better fit and reduced factors [34]. It finds the characteristics that describe the data in best possible way and looks for properties that show maximum variations across the data. Applying PCA before decision tree induction increases classification accuracy [35]. Sparse

PCA produces models using sparse loading in which subset from original variables are combined linearly to form new [36]. It isolates noise.

Use of PCA:

1) PCA is a black box that is being widely used for data analysis. It is abundantly is almost every area from computer graphics to medical as it is simple and non-parametric to deal with complex data.
2) Prediction- It is a means of extracting the patterns that dominate in the set of predictable variable sets leading to the building of model. One of the major advantage is its orthogonal nature, so no issue related to multi-collinear behavior is predicted. Even it identifies the small set of components that shows the maximum variance in data.
3) Outlier detection- Outliers can also be identified using this approach. It is needed as some variables deviates from the particular distribution for some specific reason. Removal of such data provides better assumptions and procedures to be fulfilled.
4) Classification- In preprocessing, PCA is one of the feature reduction technique which is further used classifying the data. It is not a classifier but assumes new variables that 'fit' the model. Cross-validation is applied as it prevents from accidental re-fitting of the dataset while testing.
5) Variable selection- It is a tool used for selecting the variables based on magnitude of their coefficient values. Choosing the variable of largest variance is analogous to feature extraction.

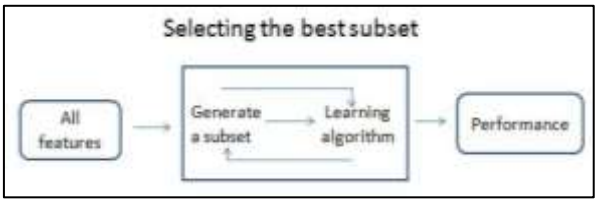


B. Factor Analysis

Factor analysis is a very useful tool for analyzing the relationship between the variables and concepts from large and complex data sets. The factor value gives overall variance to explain the variations. Factor loadings are visualized as regression coefficient. Variables are treated as independent if they have small covariance [43]. Factor analysis is done by using two techniques exploratory or confirmatory. Correlations are determined among the variables in exploratory and confirmatory already contains the thoughts that the actual had.

C. Wrapper methods

In wrapper methods, we try to use a subset of features and train a model using them. Based on the inferences that we draw from the previous model, we decide to add or remove features from your subset. The problem is essentially reduced to a search problem. These methods are usually computationally very expensive. It includes forward selection, backward elimination, recursive feature elimination.



General procedure for wrapper method

1) Forward wrapper: It is the simplest model, add suitable variables one at a time until the best model is reached.

The process starts with initially finding the AIC value of all the variables. Further, it is compared with individual AIC values of variables. The variable with lowest AIC value is taken for further processing. The process continues till we find the low values and list of variables are listed as the subset in the form of coefficients. As it is a forward wrapper, process starts with null attributes and in each step variables keep on adding.

2) Backward wrapper: It works as general model and crops variables one at a time till the best model is reached.

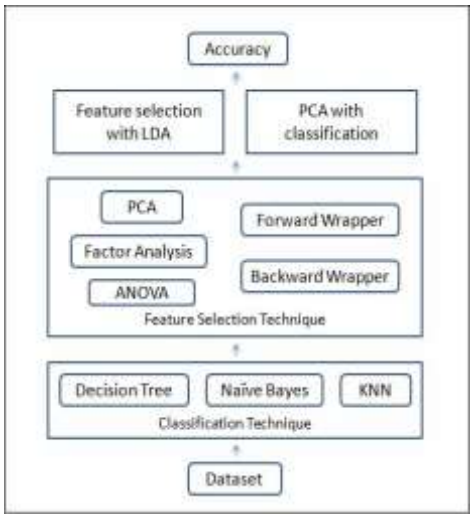
In this method, process starts same like forward method but with the list of all variables. The process runs backward, deleting the variables from the list having low value of AIC step by step. Thus, finally get the subset of subset of variables, not having values less than preferred AIC value in each step.

The process wrapper feature selection starts with randomly creating the copies of variables by shuffling them, called the shadow variables. The extended data is trained by applying classifier forest and then importance of each feature is evaluated. The process continues finding that actual features that are more significant compared to shadow variables and continuously variables are removed which are not significant. The process stops when it reaches the specified limit by removing (backward) or accepting (forward) the features.

D. Accuracy using LDA

LDA (Linear Discriminant Analysis) predicts future data based on some past data. Prior probability and means is calculated, assuming one variable as outcome. After that, coefficient of LD is created using the classes of that one significant variable chosen. Now, the new class is predicted and comparison is done between the predicted dependent variables and existing variables. For the prediction purpose, matrix is created in which columns gives the total number of rows that are actually present and diagonal gives the values correctly predicted. And finally, accuracy level of LDA is calculated, it sums up all diagonal values divided by sum of all values in matrix.

E. Proposed methodology



In the proposed methodology, try to improve accuracy of dataset by applying the combination of classification and feature selection technique. Firstly, accuracy is evaluated using decision tree, naïve bayes and k-NN, to find the best algorithm. To improve the performance of data, feature selection is applied including various filter and wrapper methods. By using this, relevant subset of attributes is obtained. To analyze its performance LDA is also applied before and after feature selection. As the PCA is proved to be the best technique with respect to arrhythmia data, it has given highest accuracy. After that, compare the result with classification algorithm. The applied methodology has given good accuracy result with feature selection on dataset and it is being discussed in the next chapter.

VII. Results and Discussion

This chapter summarizes the results by dividing it in three different sections- before feature selection, feature selection techniques and results after applying the technique.

A. Before Feature Selection

Data analysis is a process of providing the summary of the data collected for research and finding its meaning. The classification technique, decision tree is applied on Cardiac Arrhythmia Database which consists of various attributes age, gender, height, weight and other ECG attributes. The aim of using such data to analyze the situations in which there are more chances of having Cardiac Arrhythmia.

Classification in data used by many organizations for retrieving the needed information meeting its requirements and it also motivates for implementing more and more technologies for classification. It helps in risk management and even in legal discoveries. Data strategies are different from one organization to another. Imbalance in data refers to many classification problems, where representation of classes are not equal. Handling imbalance in data, we builds predictive models for such data. Predictive models are made up of attributes which are also the predictors, used for estimating the future behavior using data sets. Here, we are making estimations using a heart disease called arrhythmia. Arrhythmia is a disease in which irregular heart-rate starts occurring.

Classification technique	Accuracy

Decision tree	71.43%
Naïve Bayes	68.75%
K-NN	62.50%

Table 4.1: Accuracy measure of data set using techniques in rapid miner tool

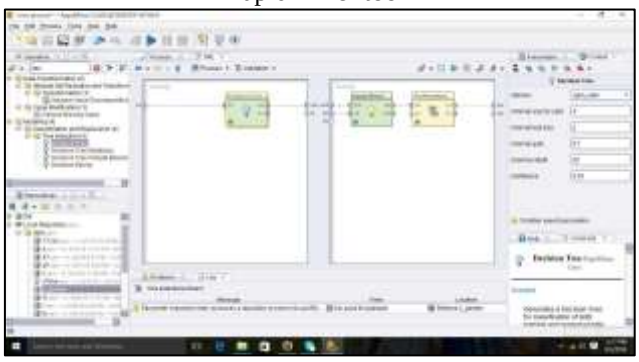


Fig. 4.1: Decision tree accuracy design

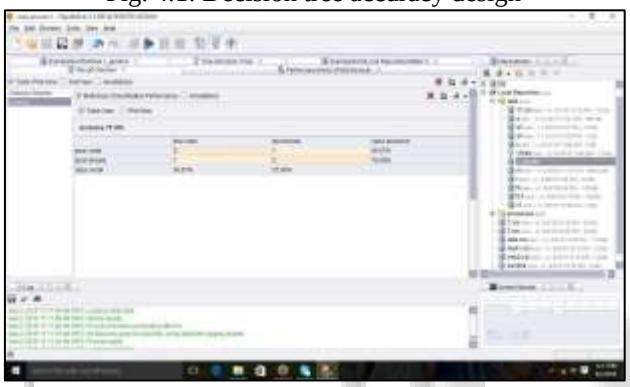


Fig. 4.2: Decision tree accuracy

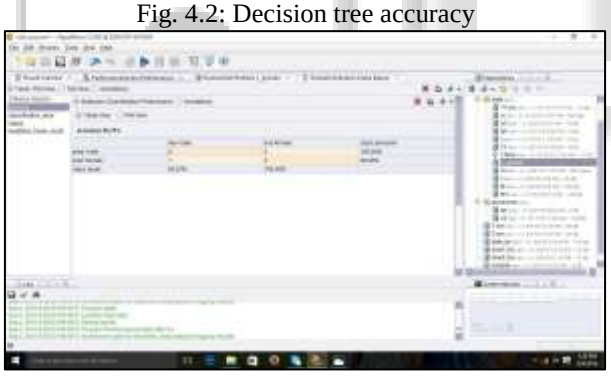


Fig. 4.3: Naïve Baye's Accuracy

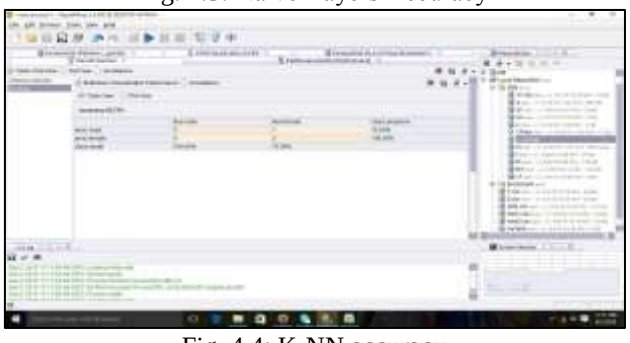


Fig. 4.4: K-NN accuracy

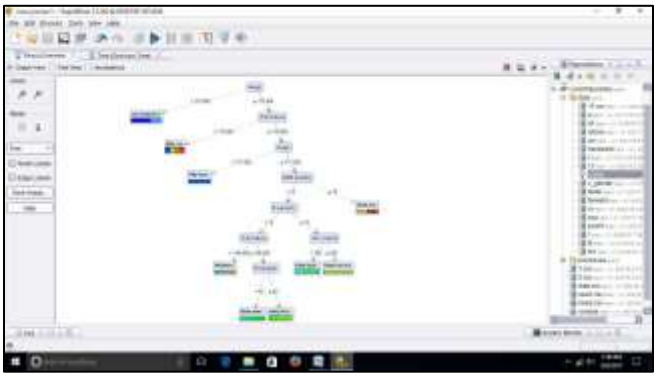


Fig. 4.5: Decision tree classification of cardiac arrhythmia training dataset

The performance of the decision tree classifier is evaluated by finding the accuracy. Here we had used rapid miner tool for this purpose. From table 4.1, we can easily evaluate that the classifier used for the data classification has the highest accuracy comparing with other two techniques. From the graphs in fig.4.6 (a) we can conclude that the most approximate age of having the chances of arrhythmia is 40-55. Major reasons are hypertension or abnormal eating habits. For this age group P-R interval ranges from 120-148.5. 80% people are having QRS values greater than 73. From (b), females(55%) are more prone to this disease and the rest are males(45%), (c) denotes 95% have QRS values 70-102 msec. For males, QRS value will be greater than 90.5 and for females, less than or equal to 90.5. P interval value for males is more than 76 and for females less than equal to 76. Height is also a major attribute in decision making, 87% people heighted less than 171.5 cm are having more chances of arrhythmia. Fig 4.6(e) denotes the pie chart of weight of person in heart-rate percent. There are different variations accordingly.

B. Feature Selection Technique

Variables	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Variable 1	13.00	44.00	47.00	47.73	54.75	75.00
Variable 2	0.00	0.00	1.00	0.54	1.00	1.00
Variable 3	70.00	77.25	82.50	86.68	91.00	138.00
Variable 4	0.00	130.5	149.5	149.3	172.2	251.0
Variable 5	321.0	354.2	376.5	372.0	385.2	401.0
Variable 6	122.0	156.2	163.0	162.9	174.0	189.0
Variable 7	39.00	66.25	80.00	85.86	93.25	183.00
Variable 8	-24.00	30.00	61.50	51.27	76.00	107.00
Variable 9	-24.00	29.50	40.50	40.32	54.75	78.00
Variable10	-17.00	45.00	56.00	48.64	68.00	78.00
Variable11	-13.00	33.25	52.50	47.23	66.00	88.00
Variable12	53.00	66.00	70.00	68.95	72.75	84.00
Variable13	0.00	0.00	0.00	1.63	0.00	20.00
Variable14	36	40	44	50	51	92
Variable15	0.00	0.00	18.00	21.82	39.00	80.00
Variable16	20.00	21.00	24.00	26.18	28.00	48.00

Table 4.2: Summary of Quartile values of the variables of dataset

Using the arrhythmia data, the analysis is done to get the results. All the variables defined have its range and we need to find their behavior with respect to decision tree and

factor analysis. Correlations indicate the linear relationships between the different features. Dependency gives the probabilistic independence which is not satisfied. Fig 4.7 shows the correlations between the various variables and their dependency. It can be positive, negative or no correlations on the basis of its characteristics. In the following figure, the variables are indicating different correlations. The plots indicate that how one value moves with respect to the other.

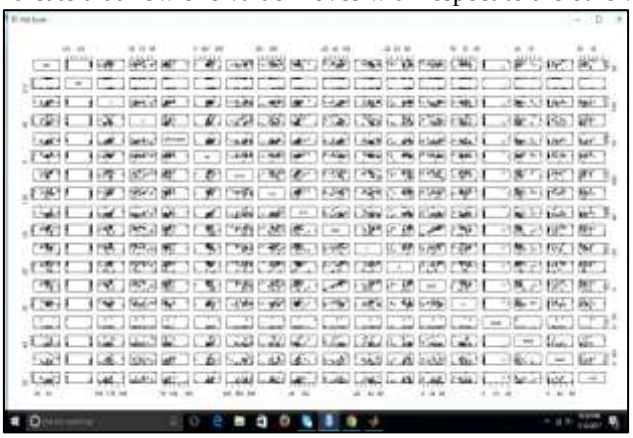


Fig. 4.7: Correlation pairs of features

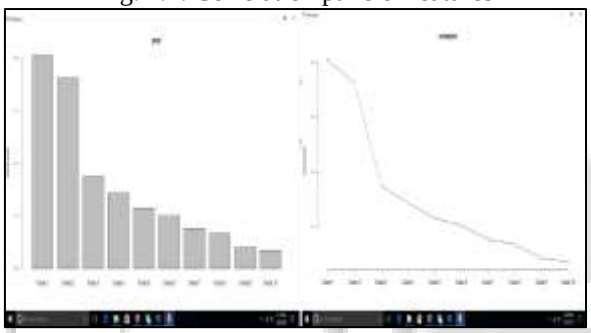


Fig. 4.8: Plot of variances of 10 components of PCA PCA

Factor analysis is done on correlations to describe variations among the observed variables. Principal Component Analysis (PCA) is the exploratory technique for factor analysis. Component values are evaluated using R. Fig 4.8 give the plot of ten component values in PCA. Fig 4.9 denotes the screeplot for the same. The figure shows that five components have value greater than one and these help in analyzing the other parameters.

No. of Components	Standard Deviation	Proportion of Variance	Cumulative Proportion
Component 1	2.0108	0.2527	0.2527
Component 2	1.9023	0.2261	0.4781
Component 3	1.3216	0.1091	0.5880
Component 4	1.1955	0.0893	0.6774
Component 5	1.0747	0.0721	0.7496
Component 6	1.0101	0.0637	0.8133
Component 7	0.8779	0.0481	0.8615
Component 8	0.8184	0.0418	0.9034
Component 9	0.6515	0.0265	0.9299
Component 10	0.5946	0.0220	0.9520
Component 11	0.5237	0.0171	0.9692
Component 12	0.4819	0.0145	0.9837
Component 13	0.3463	0.0074	0.9912
Component 14	0.3009	0.0056	0.9968
Component 15	0.2194	0.0030	0.9998
Component 16	0.0411	0.0001	1.0000

Table 4.3: Standard Deviation, variance and Cumulative Proportion Of components evaluated by PCA

Table 4.3 depicts standard deviation, variance and cumulative proportion of components that is the summary of correlations of 16 components or variables. It is the measure of dispersion. Component 16 is least dispersed. Proportion of variance that the first component explains is 25% and so on. Now we want to retain the component that has Eigen value greater than one. The cumulative proportion till comp 6 is more than 80%, these will give the most fittest values. All 16 components explain the full variation in data. Here we are looking for those variables whose possible loadings are strong on components. The values represent the correlation between components and variables.

Bi-plot in fig. 4.10 is between comp1 and comp2, they have highest value. At the centre original variables are present and represents how they are lined up on component space. Numbers denote the observations as statistics. From the observation V4, V7 has highest correlation from comp1 and V2 from comp2.

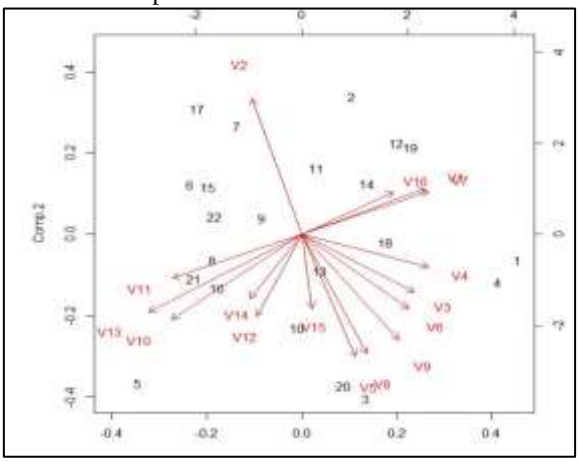


Fig. 4.10: Bi-plot between comp1 and comp2.

	Chi Square value	Degree of freedom	p-value
Factor=6	33.43	39	0.721
Factor=5	45.72	50	0.646
Factor=4	73.79	62	0.145

Factor=3	92.6	75	0.0821
Factor=2	117.6	89	0.0228
Factor=1	162.17	104	0.000229

Table 4.4: Test of hypothesis on factors by factor analysis.

```
call:
factanal(x = x, factors = 6)

Uniquenesses:
   V1    V2    V3    V4    V5    V6    V7    V8    V9   V10   V11   V12   V13   V14
0.161 0.255 0.230 0.489 0.253 0.388 0.005 0.472 0.096 0.005 0.005 0.763 0.013 0.680
  V15   V16
0.005 0.644

Loadings:
    Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
V1    0.131  -0.260  -0.796  -0.332
V2   -0.818  -0.197                 -0.173
V3    0.708  -0.228  -0.421   0.154  -0.118
V4    0.368  -0.117  -0.200  -0.104          0.554
V5    0.689   0.374  -0.199                 0.297
V6    0.492  -0.366   0.256          0.168   0.371
V7   -0.121  -0.185  -0.411  -0.461  0.468   0.588
V8    0.680          0.222
V9    0.674  -0.188   0.479  -0.235  0.292   0.211
V10           0.965   0.204
V11  -0.130   0.192   0.202   0.932  0.112  -0.132
V12   0.237   0.254           0.336
V13           0.900   0.205   0.331         -0.157
V14                   0.539                 -0.105
V15   0.192           0.116          0.965
V16          -0.365           -0.115 -0.275  0.358

                Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
SS loadings       3.099   2.497   1.838   1.555   1.418   1.129
Proportion Var    0.194   0.156   0.115   0.097   0.089   0.071
Cumulative Var    0.194   0.350   0.465   0.562   0.650   0.721
```

Fig. 4.11: Uniqueness, loadings and other parameters by factor analysis at factor=6

Fig. 4.12: Result of forward wrapper

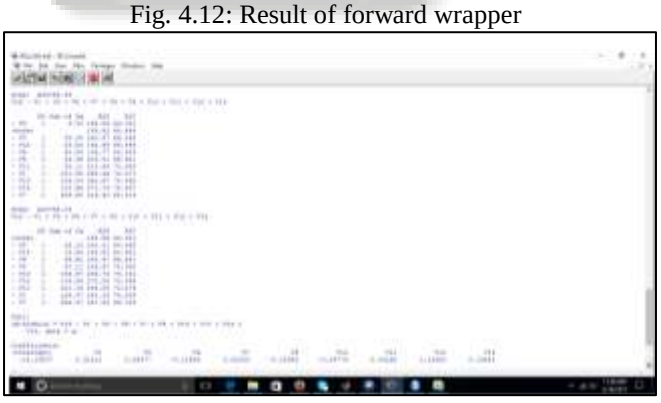


Fig. 4.13: Result of backward wrapper

Instead of talking about what percentage of variance is explained, in fig 4.1 uniqueness of factors is considered. V12, V4 has highest unique values. If the uniqueness is high, cumulative is low which means variations explained by these is low on these. Considering factor loadings, V5 and V8 of factor1 is high and inverse loading is on V2, V7, V12 so on. The chi-square value gives significant relationship between two variables, observed and expected. It depends on the degree of freedom. The p-value in it denotes the probability when chi-square value is large and there is no relationship between the factors. The observed correlations are significant when the p-value is less than 0.05, so factor1 and factor2 gives the significant values.

VIII. ANOVA

ANOVA stands for analysis of variance. It is a statistical analysis used to test the degree that how two groups of variables vary. Variance (difference) gives the significant results from the experiment performed. Hypothesis is assumed in this process. Null hypothesis assumes no difference between groups. In alternate hypothesis, there is difference between groups and ANOVA is performed. P value is considered as an important analysis criteria. The cut off value is 0.05. P-value is less than 0.05, than significant result and null-hypothesis is rejected. If p-value is greater than 0.05, it is not considered as significant.

Variables	Gender	QRS duration	PR interval	QT interval	T interval	P interval	QRS	T	PQRST	Heart-rate	Q wave	R Wave	S wave	NOID
Age				0.00561			0.0416	0.013	0.0156					
Gender		0.00265	0.0426		0.0123	0.0128							0.0489	
QRS duration					0.0161								0.00051	
PR interval						0.00141								0.0484
QT interval								0.00702	0.0236					
T interval						0.00182							0.0263	
P interval											0.0478			0.00168
QRS									5.26e-12					
T									0.00834					
P														
QRST														
Heart-rate														
Q wave														
R wave													0.0354	0.0431
S wave														

Table 4.5: P values of variables using ANOVA

A. LDA

Linear discriminate analysis is based on some past data we try to predict future data. All input or independent variables are continuous and output is categorical in nature. The whole analysis is done using R. The outcome expected from LDA is age variable. Probability counts each class in data. After finding the probabilities, prediction is going to take place using the same set of input data, we try to predict dependent variables and then comparing the predicted dependent variables with the existing variables. The linear combination of coefficient (scaling) is for each linear discriminate. In this case, we have four linear discriminate. Single value discriminate gives the ratio of between and within group standard deviation on linear discriminate variables. There is a need to find out the values which have been predicted for ages. Using R, ldaclass stores the information predicted. Using ldatable, columns gives the total number of rows that are actually present. The diagonal gives the values that are correctly predicted. Accuracy level of LDA is calculated. It sums up all diagonal values divided by sum of all values in matrix. So, it gives 54.5%. The model LDA using R gives the accuracy level of 54.5% for the results which are already there and predict from arrhythmia dataset.

LDA is an analysis method which evaluates the support that a prior group of variables provide the predicted group of variables. It is closely related to PCA, ANOVA, FA, linear regression. It is a method of finding linear combination of variables or features that successfully separates two or more classes of objects. PCA is an unsupervised learning technique (don't use class information) while LDA is a supervised (uses class information).

1) Both reduce dimensions.

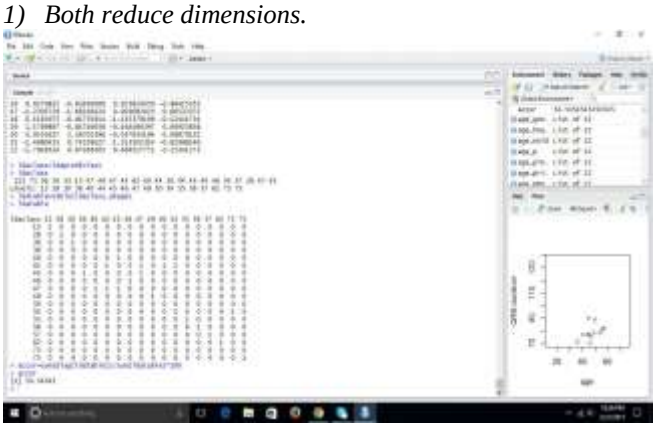


Fig. 4.14: Accuracy using LDA

2) After Feature Selection

In the previous section, subset of attribute is selected using various feature selection techniques of filter and wrapper methods with accuracy. Table 4.6 shows the LDA accuracy with different class of arrhythmia. The accuracy is evaluated before and after the relevant selection of attributes. Forward and backward wrapper methods are used for this purpose. The accuracies calculated in table using LDA before feature selection is much low. The Coronary class of arrhythmia has the highest LDA accuracy with 98% and lowest 37.5%, without feature selection. Improving it, forward and backward wrapper is applied and accuracies jump high. Forward wrapper gives better accuracies with respect to backward highest with normal arrhythmia 90%.

Classes	LDA	LDA after Forward Wrapper Feature Selection	LDA after Backward Wrapper Feature Selection
Normal Arrhythmia	54.54%	90%	75%
Coronary	98%	-	-
Sinus Tachycardy	61.53%	89.23%	-
Sinus Bradycardy	57.9%	43.83%	11%
RBBB	97.33%	-	-
Others	77.27%	81.81%	86.36%

Table. 4.6: LDA of different classes of arrhythmia and accuracy after performing forward and backward feature selection method

For the further processing, the data of normal arrhythmia is considered. Using different feature selection methods, subset of variables are chosen. As shown in table, the methods have given results with different number of variables. LDA accuracy after feature selection has shown high accuracy when compared with earlier results with highest 95.45% (PCA).

Variables more than 50% priority or used in more than 3 feature selection algorithms are listed in table 9. P-interval, QRS, QRS duration, T, P, QRST are the six variables has shown more relevance with respect to other ten variables.

Feature selection Method	Total no of features	No. of features selected	LDA
Backward wrapper	16	9	75%
Forward wrapper	16	8	90%
ANOVA	16	7	90.90%
Factor analysis	16	6	90.90%
PCA	**16**	**3**	**95.45%**

Table 4.7: Comparison of LDA accuracies with respect to number of features selected and different feature selection methods.

Variables	PCA	Factor	ANOVA	Backward	Forward
P interval	□	□	□		
QRS	□		□	□	
QRS duration		□	□	□	
T		□	□		□
P		□		□	□
QRST		□		□	□

Table 4.8: Variables more than 60% or used in more than 3 feature selection algorithms

No of components	LDA
2	54.54%
3	95.45%
4	77.27%
5	63.63%
6	54.54%
7	50%
8	50%

Table 4.9: Analysis of number of components by PCA and highest accuracy with 3 components.

In table 4.7, accuracies are evaluated applying different feature selection methods including filter (ANOVA, factor analysis, PCA) and wrapper methods (forward and backward). With less number of attributes, more accurate results are seen. Comparing both feature selection methods, we came to the conclusion that filter methods gives more appropriate results with less features. Wrapper methods make the model more over-fitted. PCA method, with 3 components of features gives 95.45% accuracy. Table 4.9 has listed the LDA accuracies calculated for the components formed from PCA method and has maximum accuracy with 3 components.

After all, accuracy of PCA selected subset is calculated using classification techniques of decision tree, naïve bayes in table 4.10. The three different categories are used for this purpose- training/testing, cross validation and split criteria. With training and testing, AD tree and random forest has shown highest accuracy. The different methods has shown high accuracy with different cross-validation and split percentages. Using cross validation, J48 and with split, random tree has given highest accuracies.

Techniques	Training/Testing (%)	Cross-validated (%)	Split (%)
AD Tree	98	72.72 (8)	90.90 (50)
BF Tree	86.36	72.72 (20)	75 (65)
FT	54.54	54.54	55.5
J48	95.45	77.27 (10)	84.61 (40)
LAD Tree	94	68.18 (10)	84 (40)
Random Forest	98	77 (8)	91.66 (45)
Random Tree	91	77 (12)	92.30 (40)
REP Tree	77	63.63 (12)	69.23 (40)
Simple Cart	77	68.18 (10)	75 (65)
Naïve Bayes	86.36	72 (8)	75 (10)

Table 4.10: Accuracy results after feature selection using different classification methods

IX. ROC

ROC is a receiver operating characteristic curve where sensitivity denotes true positive on y axis and specificity is true negative on x axis. If the curve follows the left hand border of true positive then the top border of ROC curve, the more accurate is the test. On the other hand, if the curve is closer to 45 degree diagonal of ROC, less accurate is test. Here, plotting is done on various threshold settings. Sensitivity measures the proportion of positives that are truly classified where as specificity measures the proportion of negatives that are truly classified in machine learning.

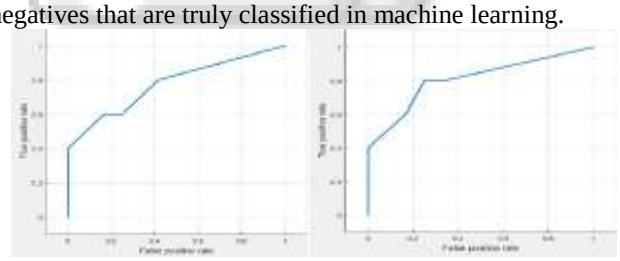


Fig. 4.15: ANOVA Fig. 4.16: Backward wrapper

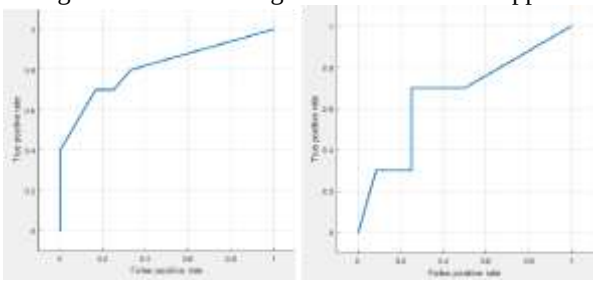


Fig. 4.17: Factor

S.No	Technique	ROC accuracy (%)
1	Backward	80
2	ANOVA	77
3	Factor	81
4	Forward wrapper	66
5	PCA	90.8

Table 4.11: ROC accuracy for arrhythmia after feature selection using matlab

It is generated by plotting cumulative distribution function. It shows relative behavior as it is comparison between operating characteristics of sensitivity and specificity. Considering ROC curve, PCA has shown highest accuracy with 90.8%, after that factor with 81% with positive class male and least with forward wrapper.

X. CONCLUSION

Classification techniques are used to analyze the data to categorize them in classes. Huge amount of data has become a very typical task. Classification techniques are used to solve this problem at a great extent. The performance of the decision tree classifier is evaluated by finding the accuracy. Here we had used rapid miner tool for this purpose.

Feature selection algorithm shows complex behavior with respect to data size, labels and the rules from features. In the presence of high dimensions, little correlation is often seen between errors for selected and best feature sets. Keeping feature set small, more accurate error estimations are made. Visualization of data is very important. Working with subsets of data will give additional insight. Factors are identified using various feature selection algorithm to identify the correlation between different variables.

1) Decision classifier used for the data classification has the highest accuracy with 71.3%, comparing with other two techniques naïve bayes and k-NN before feature selection (table 4.1).
2) The most approximate age of having the chances of arrhythmia is 40-55. Females (55%) are more prone to this disease and the rest are males (45%), 95% have QRS values 70-102 msec (fig 4.6).
3) Linear Discriminate Analysis is used as one of the parameter for finding the best accuracy of 95.45% using PCA with 3 components (table 4.9).
4) With training and testing, AD tree and random forest has shown highest accuracy. Using cross validation, J48 and with split, random tree has given highest accuracies.
5) Considering ROC curve, PCA has shown highest accuracy with 90.8%, after that factor with 81% with positive class male and least with forward wrapper (table 4.11).

After all, accuracy of PCA selected subset is calculated using classification techniques of decision tree, naïve bayes and k-NN. The three different categories are used for this purpose-training/testing, cross validation and split criteria. The different methods has shown high accuracy with different cross-validation and split percentages. ROC is generated by plotting cumulative distribution function. It shows relative behavior as it is comparison between operating characteristics of sensitivity and specificity.

A. Limitations:

The analysis using PCA relies totally on linear assumptions. The original features of using this technique shows orthogonal behavior. If the attributes of data shows non-linear relationships, some dimensions are added to it. Addition process increases the computational complexity and gives

unstable solutions. PCA is scale variant, if we change the scale of attributes used, it will show different results after applying.

XI. FUTURE USE

Future work includes many more additional experiments on analyzing the data from other fields. Research based on hybrid approach of feature selection techniques and their impact on classification. Handling the missing values using various techniques. Search strategies using embedded feature selection technique. Integration of other sources of information, different combination of feature selection methods and fuzzy set theory to improve the model development. Analyzing the results with unsupervised technique. Hybrid model can also be used to improve and can get better results.

REFERENCES

[1] Saini I. and Saini B., "Cardiac Arrhythmia Classification Using Error Back Propagation Method", International Journal of Computer Theory and Engineering, Vol. 4, No. 3, June 2012.
[2] Elsayyad A., Nassef A., Baareh A., "Cardiac Arrhythmia Classification Using Boosted Decision Trees", International Review on Computers and Software (I.RE.CO.S.), Vol. 10, N. 3 ISSN 1828-6003 March 2015.
[3] Kohli N. and Verma N., "Arrhythmia classification using SVM with selected features", International Journal of Engineering, Science and Technology Vol. 3, No. 8, 2011.
[4] Choudhary K., Bajaj P. , "Automated Prediction of RCT (Root Canal Treatment) Using Data Mining Techniques: ICT in Health Care", International Conference on Information and Communication Technologies (ICICT 2014).
[5] Bellaachia A., Guven E., "Predicting Breast Cancer Survivability Using Data Mining Techniques".
[6] Parvathi I., Rautaray S., "Survey on Data Mining Techniques for the Diagnosis of Diseases in Medical Domain", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014.
[7] Singh R. ,Sivabalakrishnan M., "Feature Selection of Gene Expression Data for Cancer Classification: A Review" , 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15).
[8] Aneeshkumar A., Venkateswaran C., "Reverse sequential covering algorithm for medical Data mining", Procedia Computer Science 47 (2015).
[9] Sasikalaa S., Balamuruganb S., Geetha S., "A Novel Feature Selection Technique for Improved Survivability Diagnosis of Breast Cancer", 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15).
[10] Ilayaraja M., Meyyappan T., "Efficient Data Mining Method to Predict the Risk of Heart Diseases through Frequent Itemsets", 4th International Conference on Eco-friendly Computing and Communication Systems, ICECCS 2015.
[11] Nagarajana G., Minub R., Muthukumar B., Vedanarayanan V., Sundarsingh S.,"Hybrid Genetic Algorithm for Medical Image Feature Extraction and selection", International Conference on Computational Modeling and Security (CMS 2016).
[12] Baitharua T., Kumar S., "Analysis of Data Mining Techniques For Healthcare Decision Support System Using Liver Disorder Dataset", International Conference on Computational Modeling and Security (CMS 2016).
[13] Kaur S. and Bawa R., " Future Trends of Data Mining in Predicting the Various Diseases in Medical Healthcare System", International Journal of Energy,
[14] Ahmad P., Qamar S., Rizvi S., "Techniques of Data Mining In Healthcare: A Review", International Journal of Computer Applications (0975 – 8887) Volume 120– No.15, June 2015.
[15] Bekri F., Govardhan A., "Association of Data Mining and Healthcare Domain: Issues and Current State of the Art", Global Journal of Computer Science and Technology Volume 11 Issue 21 Version 1.0 December 2011.
[16] Anthimopoulo M., Christodoulidis S., Ebner L., Christe,A., Mougiakakou S., "Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network", Ieee Transactions On Medical Imaging, Vol. 35, No. 5, May 2016.
[17] Bellazzi R., Zupan B., "Predictive data mining in clinical medicine: Current issues and guidelines", international journal of medical informatics 7 7 (2 0 0 8).
[18] Adya M., "Data Mining in Health-Care: Issues and a Research Agenda", AMCIS 2000 Proceedings, 2000.
[19] Durairaj M., Ranjani V., "Data Mining Applications In Healthcare Sector: A Study", International Journal Of Scientific & Technology Research Volume 2, Issue 10, October 2013.
[20] Khaleel M., Pradham S., Dash G., " A Survey of Data Mining Techniques on Medical Data for Finding Locally Frequent Diseases", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue
[21] Bhukya D. and Ramachandram S., "Decision Tree Induction: An Approach for Data Classification Using AVL-Tree", International Journal of Computer and Electrical Engineering, Vol. 2, No. 4, August, 2010 1793-8163.
[22] Madadipouya K., "A New Decision Tree Method For Data Mining In Medicine", Advanced Computational Intelligence: An International Journal (ACII), Vol.2, No.3, July 2015.
[23] Teli S., Kanikar P., " A Survey on Decision Tree Based Approaches in Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, 2015.
[24] Țăranu I., "Data mining in healthcare: decision making and precision", Database Systems Journal vol. VI, no. 4/2015.
[25] Dey M., Rautaray S., "Study and Analysis of Data mining Algorithms for Healthcare Decision Support System", International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, 470-477.

[26] Goel V., "Decision analysis: applications and limitations", CAN MED ASSOC J 1992.
[27] Werner E., Wheeler S. and Burd I., "Creating Decision Trees to Assess Cost-Effectiveness in Clinical Research", J Biomet Biostat 2012.
[28] Elwyn G., Edwards A., Eccles M., Rovner D., "Decision Analysis In Patient Care", The Lancet , Vol 358 , August 18, 2001.
[29] Soleimanian F., Mohammadi P., Hakimi P., "Application of Decision Tree Algorithm for Data Mining in Healthcare Operations: A Case Study", International Journal of Computer Applications (0975 – 8887) Volume 52 – No. 6, August 2012.
[30] Quinlan J., "Induction Of Decision Trees", Machine Learning 1: 81-106, 1986.
[31] Szolovits P., "Uncertainty and Decisions in Medical Informatics", Methods of Information in Medicine, 34:111–21, 1995.
[32] Teli S., Kanikar P. ," A Survey on Decision Tree Based Approaches in Data Mining",
[33] International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, 2015.
[34] Srimani P.K. And Koti M.S, "Evaluation Of Principal Components Analysis (Pca) And Data Clustering Techniques (DCT) On Medical Data", International Journal Of Knowledge Engineering, ISSN: 0976-5816, Volume 3, Issue 2, 2012.
[35] Sabharwal C., Anjum B., "Principal Component Analysis as an Integral Part of Data Mining in Health Informatics", Proceedings of 31st International Society Conference on Computers And Their Applications CATA 2016, pp. 251-256, April 05, 2016.
[36] Wimmer H., Powell L., "Principle Component Analysis for Feature Reduction and Data Preprocessing in Data Science", 2016 Proceedings of the Conference on Information Systems Applied Research Las Vegas, Nevada USA.
[37] Sjostrand K., Stegmann M. and Larsen R., "Sparse Principal Component Analysis in Medical Shape Modeling".
[38] Jayaprada S., "Enhanced C-Means Clustering with PCA for medical dataset",IJDCST, April-May-2016.
[39] Luukka P., "A New Nonlinear Fuzzy Robust PCA Algorithm and Similarity Classifier in Classification of Medical Data Sets", International Journal of Fuzzy Systems, Vol. 13, No. 3, September 2011.
[40] Kalaiselvi.R, Premadevi.P and Hamsathvani.M, "Weighted Principle Component Analysis for Dimensionality Reduction in Medical Dataset", IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 5, May 2015.
[41] Salama G., Abdelhalim M., and Zeid M., "Breast Cancer Diagnosis on Three Different Datasets Using Multi-Classifiers", International Journal of Computer and Information Technology (2277 – 0764) Volume 01– Issue 01, September 2012.
[42] Naik G., Selvan S., Gobbo M., Acharyya A., Nguyen H., "Principal Component Analysis Applied To Surface Electromyography: A Comprehensive Review", IEEE Access, VOLUME 4, 2016.
[43] Hu B., Dai Y., Su Y., Moore P., Zhang X., Mao C., Chen J., Xu L., "Feature Selection for Optimized High dimensional Biomedical Data Using An Improved Shuffled Frog Leaping Algorithm", 2016 Ieee.
[44] Williams B., Onsman A., Brown T., " Exploratory factor analysis: A five-step guide for novices", Journal of Emergency Primary Health Care (JEPHC), Vol. 8, Issue 3, 2010.
[45] Pinto C., "Data Reduction I: PCA and Factor Analysis", Data Analysis Seminars 11November 2009.
[46] Anna B. and Jason W., "Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis", Practical Assessment, Research & Evaluation, Volume 10 Number 7, July 2005.
[47] Soman T., Patrick O., "Classification of Arrhythmia Using Machine Learning Techniques".
[48] Batra A., Jawa V., "Classification Of Arrhythmia Using Conjunction Of Machine Learning Algorithms And ECG Diagnostic Criteria", International Journal Of Biology And Biomedicine, Volume 1, 2016.
[49] Vishwa A., Lal M., Dixit S., Vardwaj P., "Clasification Of Arrhythmic ECG Data Using Machine Learning Techniques", International Journal of Artificial Intelligence and Interactive Multimedia, Vol. 1, Nº 4, 2011.
[50] Kirtania R., Mali K., " Cardiac Arrhythmia Classification using Optimal Feature Selection and K-Nearest Neighbour Classifier", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 1, January 2015.
[51] Samad S., Khan S., Haq A., Riaz A.,"Classification of Arrhythmia", International Journal of Electrical Energy, Vol. 2, No. 1, March 2014.
[52] Bhardwaj P., Choudhary R., Dayama R., "Analysis and Classification of Cardiac Arrhythmia using ECG Signals", International Journal of Computer Applications (0975 – 8887) Volume 38– No.1, January 2012.
[53] Priyadharshini V., kumar S., "An Enhanced Approach on ECG Data Analysis using Improvised Genetic Algorithm", International Research Journal of Engineering and Technology (IRJET), Volume: 02 Issue: 05 | Aug-2015.
[54] Dhakate P., Rajeswari K., Abin D., " Analysis of Different Classifiers for Medical Dataset using Various Measures", International Journal of Computer Applications
[55] Rondina J., Hahn T. et. all, "SCoRS—A Method Based on Stability for Feature Selection and Apping in Neuroimaging", IEEE Transactions On Medical Imaging, Vol. 33, No. 1, January 2014.
[56] Huda S., Yearwood J. et. all, "A Hybrid Feature Selection With Ensemble Classification for Imbalanced Healthcare Data: A Case Study for Brain Tumor Diagnosis", IEEE Access, Volume 4, 2016.
[57] Meng X. , Huang Y., Rao D. , Zhang Q., Liu Q., "Comparison of three data mining models for predicting diabetes or prediabetes by risk factors", Kaohsiung Journal of Medical Sciences (2013) 29, 93-99.

[58] Kandhasamy J., Balamurali S., "Performance Analysis of Classifier Models to Predict Diabetes Mellitus", Procedia Computer Science 47 (2015) 45 – 51.
[59] Kaur R., Verma P., "Improved MLP-NN based approach for Lung Diseases Classification", International Journal of Computer Applications (0975 – 8887) Volume 131 – No.6, December2015.
[60] Gokilam G., Shanthi K., "Performance Analysis of Various Data mining Classification Algorithms on Diabetes Heart dataset", COMPUSOFT, An international journal of advanced computer technology, 5 (3), March - 2016 (Volume-V, Issue-III)
[61] Boukenze B., Mousannif H. and Haqiq A., "Performance Of Data Mining Techniques to Predict In Healthcare Case Study: Chronic Kidney Failure Disease ", International Journal of Database Management Systems (IJDMS) Vol.8, No.3, June 2016.
[62] Thangaraju P., Barkavi G., Karthikeyan T., "Mining Lung Cancer Data for Smokers and Non-Smokers by Using Data Mining Techniques ", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 7, July 2014.
[63] Swathi P., Vital P., "Statistical Analysis and Predicting Kidney Diseases using Machine Learning Algorithms", International Journal of Engineering Research & Technology (IJERT) Vol. 4 Issue 07, July-2015.
[64] Jena L., Kamila N., "Distributed Data Mining Classification Algorithms for Prediction of Chronic-Kidney-Disease", International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-4, Issue-11), 2015.
[65] Kohli N. and Verma N., "Arrhythmia classification using SVM with selected features", International Journal of Engineering, Science and Technology, Vol. 3, No. 8, 2011, pp. 122-131.
[66] Muntean M., Valean H., Cabulea L., "Feature Selection Methods For Multidimensional Datasets", International Conference on Theory and Applications in
[67] Mathematics and Informatics 17^{th}-20^{th} of September, 2015, pp. 167-177.
[68] Macedo D., Nasser Matos S., Borges H., Francisco A., " The Use Of Attribute Selection In The Banking Sector In Order To Obtain Knowledge Of Customers".
[69] Ramaswami M. and Bhaskaran R., "A Study on Feature Selection Techniques in Educational Data Mining", Journal Of Computing, Volume 1, Issue 1, December 2009, ISSN: 2151-9617.
[70] Abbas S., "Deposit subscribe Prediction using Data Mining Techniques based Real Marketing Dataset", International Journal of Computer Applications (0975 – 8887) Volume 110 – No. 3, January 2015.
[71] Yezus A., " Predicting outcome of soccer matches using machine learning", Term paper, Saint-Petersburg State University, Mathematics and Mechanics Faculty, 2014.
[72] Devaraj S. and Paulraj S., "An Efficient Feature Subset Selection Algorithm for Classification of Multidimensional Dataset", The Scientific World Journal Volume 2015, Article ID 821798,2015.
[73] Chakraborty D. And Maulik U,, "Identifying Cancer Biomarkers From Microarray Data Using Feature Selection And Semisupervised Learning", Ieee Journal Of Translational Engineering in Health And Medicine, 2014.
[74] Ajay, Venkatesh A., Jacob S., "Prediction of Credit-Card Defaulters: A Comparative Study on Performance of Classifiers", International Journal of Computer Applications (0975 – 8887) Volume 145 – No.7, July 2016.
[75] Pitt E. and Nayak R., "The Use of Various Data Mining and Feature Selection Methods in the Analysis of a Population Survey Dataset".
[76] Sheena, Kumar K., Kumar G., "Analysis of Feature Selection Techniques: A Data Mining Approach", International Journal of Computer Applications,4th International Conference on Engineering & Technology (ICAET 2016).
[77] Lavanya D., Rani U., "Analysis Of Feature Selection With Classfication: Breast Cancer Datasets", Indian Journal of Computer Science and Engineering (IJCSE), Vol. 2 No. 5 Oct-Nov 2011.
[78] Chitra K., Subashini B., "Data Mining Techniques and its Applications in Banking Sector", International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 8, August 2013.
[79] Petre R., "Data Mining Solutions for the Business Environment", Database Systems Journal vol. IV, no. 4/2013.
[80] Goswami S., Chakrabarti A., "Feature Selection: A Practitioner View", I.J. Information Technology and Computer Science, 2014, 11, 66-77.
[81] Haghighat M., Rastegari H. and Nourafza N., "A Review of Data Mining Techniques for Result Prediction in Sports", Advances in Computer Science: an International Journal, Vol. 2, Issue 5, No.6 , November 2013.
[82] Bhardwaj A., Sharma A., Shrivastava V., "Data Mining Techniques and TheirImplementation in Blood Bank Sector –A Review", International Journal of Engineering Research and Applications (IJERA), Vol. 2, Issue4, July-August 2012.

