ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 3, Issue 4, April 2018

Prediction of Heart Disease Using Hybrid Algorithm

[1] V.C.N.D.Sindhu sha, [2] R.Sai Shankar, [3] M.Sribala
[1] Student Dept of CSE, LBRCE,
[2] Student Dept of CSE, LBRCE,
[3] Sr.Asst.Professor,Dept of CSE,LBRCE.

***Abstract*- Classification is normally supervised learning method and helps to classify the data into classes and design models, where the attribute of class plays an important role for the construction of the classifier. Nearest neighbor (KNN) is more easy and simple, effective, efficient and popular algorithm for classification. KNN is a straight forward classifier and classification is done based on its nearest neighbour. For example clinical databases like that. Classification methods produce inaccurate results; hence this algorithm helps to produce accurate results for this. In India Heart disease is one of the leading cause.is 32% and in Canada it is 35%.By keeping in the view of all these considerations a perfect system that is highly decision supportive is required. Here we proposed a hybrid algorithm which is a combination of evolutionary algorithm Genetic algorithm in combination with KNN for good accuracy and perfect classification.Genetic Algorithms are used perform global search for fitness solutions and KNN used to classify them.**

***Keywords*: Data mining, Genetic algorithm, Heart disease, KNN algorithm, classification, accuracy etc.**

I. INTRODUCTION

Data mining is a process for extracting useful information from a larger data set of any raw database. It used to analyse data patterns in different sets of data using various tools. It has wide range of applications in different fields, like science and engineering. This helps in a way to provide solutions for business also to take better decisions. It has 3 phases like effective data collection and warehousing as well as computer processing. For clustering and probability for future events it implements best machine learning algorithms. Data mining is also represented as Knowledge Discovery in Data (KDD).

Classification is used for classifying data into different classes and to satisfy certain constraints. Different classification algorithms including C4.5, ID3, k-nearest neighbor classifier, Naive Bayes, SVM, and ANN are used for classification. Normally a classification technique follows three approaches Statistical, Machine Learning and Neural Network for classification. Classification methods present in data mining normally are able to process the data on large scale. It also helps us predict categorical labels of a given class further classifies the data on the basis of training set and on lables of data. It also used for classification for data which is new. Classification technique is recognized as a method for repeatedly making such decisions to handle new kind of problems. Deriving a classification procedure from a data set for which the exact classes are known priori is called as pattern recognition or supervised learning. Various examples include diagnosis of a patient, predicting stock market values and produce powerful decision support systems. The below are three main research methods that can be identified as statistical, machine learning and neural network. The nearest neighbour concept represents the classification of data point which is unknown based closest neighbour in which the class is known priorly. In KNN the nearest neighbour is computed according to the estimation of k which refers nearest neighbours are to be considered to characterize class of a sample data point. It helps in the utilization of the several closest neighbour to decide the class in where this data point falls under and also called as KNN. These are the data elements that are needed for the memory for run time and these are represented as memory-oriented methods. T. Bailey and A. K. Jain enhanced the K Nearest neighbour. Here training data is assigned to weights according to their distances from sample data. But at one particular time the memory requirements and computational complexity are same in the primary concern. To clear the problem of memory limitation the data set size is reduced. For the patterns which are repeated and cannot include extra data is also cleared from training data set. To further enhance the information focuses which don't influence the result are additionally eliminated from training data set. The nearest neighbour training data set is organized by utilizing several systems for enhancing for memory limit of K-Nearest Neighbour algorithm. The implementation of KNN is done by using 1.ball tree, 2.k-d tree, 3.nearest feature line (NFL), 4.principal axis search tree and 5.orthogonal search tree. The training data which is in tree structured form is having parts as nodes ,techniques like NFL and tunable metric which can divide the training data set based on planes. By implementing the above algorithm the KNN

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 3, Issue 4, April 2018

algorithm speed can be increased. Let us consider that an object is sampled with a set of different attributes. By considering its group is decided from its attributes; several algorithms are used to automate the process of classification. The pseudo code for the k-nearest neighbor classification algorithm is given below:

K ? number of nearest neighbors
For each object X in the test set do
calculate the distance D(X,Y) between X and every object Y in the training set
neighborhood ? the k neighbors in the training set closest to X
X.class ? SelectClass (neighborhood)
End forKNN is used in many applications such as
1) classification and interpretation
2) problem solving
3) function learning
and

4)teaching and training.

Genetic algorithm

Evolutionary computing is initiated by considering different ideas from biological science methods and to computer science methods. Genetic algorithms are very important methods for evolutionary computing methods. Evolutionary algorithms are mainly utilized for Optimization techniques.In order to solve such a type of problems, these algorithms requires a special data structure for representing and evaluating the solution from the pool of existing solutions. Evolutionary algorithms broad application area of research, those are as follows.

a) Machine intelligence
b) Travelling sales person problem
c) Expert system
d) Medicine
e) Engineering application
f) Wired and wireless communication systems

Genetic algorithms was initiated by john Holland in the year of 1975.Genetic algorithms can be used for searching problems and for optimization problems.Genetic Algorithms can be used its genetics as it's model and also for problem solving method. Each and every solution given in genetic algorithm is naturally represented through a set of chromosomes. Chromosomes are made from genes and genes are considered as individual elements (alleles) that represent the particular problem. The set of all these chromosomes is collectively called as population. Generally there are three most popular operators that are used in GA.

1) Selection:

This selection operator can be used for selecting individuals to reproduce new set of individuals. Different selection methods are listed as
Roulettewheel selection method
Random selection method
Rank selection method
Tournament selection method
Boltzmann selection method.

2) Crossover: This operator is used as the process for mixing two most highly fitted parent chromosomes and to give a birth to a new one. This particular operator is used for creating a better string.
Different methods for crossover operators are
Single point crossover method
Crossover with reduced surrogate method
Ordered crossover method
Partially matched crossover method.

3) Mutation: This mutation operator can be used for altering the new solutions of the search and to produce a best solution. Mutation generally prevents the Genetic Algorithms which are to be stocked in a localminimum of statespace.

4) Fitness function: fitness function used for genetic algorithms are the value of a given objective function for its phenotype. There are several uses for genetic algorithm. Some of its uses are given below as follows.
Its solution space is wider.
It is easy to discover global optimum in GA.
It mainly uses function evaluations.
It can efficiently handle noisy functions in a good manner.

Some of the limitations for genetic algorithms are listed below as follows.
1) It is highly difficult to Identify a suitable fitness function.
2)Normally genetic algorithms require large number of fitness evaluations
3) For identifying local optima these are not suggestible.
4) There is no method for straight forward configuration for genetic algorithms.

Heart disease

Normally there results no Heart disease whenever the artery usually gives oxygen and blood to the heart blocked completely or it is narrowed. These heart troubles occurs normally while the birth or later in the life. Based on the survey carried by register general of India, heart troubles are considered as the important reason death in India and in Andhra Pradesh.The following are different heart diseases are
Coronary heartdisease
Valvular heartdisease

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 3, Issue 4, April 2018

Hypertensive heartdisease
Ischaemic heartdisease
Cardiomyopathy
Cardiovascular disease
Inflammatory heartdisease
Heart failure

The risk factors that are common for heart trouble include more blood pressure,Abnormal blood lipids,heavy use of tobacco,Obesity inactivity in physical life,Diabetes,Age,Gender and finally the history of the Family.According to the conducted survey of WHO, at every 10 deaths 8 are due to cardio vascular heart diseases and due to diabetes. The death rate of heart disease in Andhra Pradesh is given as 30 percent. There are more preventive techniques are needed to overcome the factors for risk is highly important to overcome and handle such chronic situations of heart troubles.

The new method we proposed is a hybrid which is a mixture of genetic algorithm and KNN for improving the accuracy of classification for dataset of heart disease. Firstly we applied genetic search used to reduce highly irrelevant attributes and redundant attributes, and provide ranking to them to play a major role in the classification of heart disease. Highly ranked attributes are retained and remaining are removed further the classification algorithm applied on these retained and evaluated attributes. The generated classifier is now trained for classifying on heart disease data set which is classified as healthy or sick. The newly proposed algorithm divided into 2 major parts.

1) Firstly it implements genetic search for evaluating attributes.

2) Secondly build an accurate for the newly proposed smethod.

Step 1) data set to be loaded first.

Step 2) Now on that data set we will implement our genetic search.

Step 3) Now we can rank the attributes based on their values for this we used Roulette fitness function.

Step 4) Now form a group of subset of attributes with high ranking.

Step 5) No we can apply KNN on the set of attributes.

Now This step helps us to build the classifier and step 6 records the accuracy of the classifier.

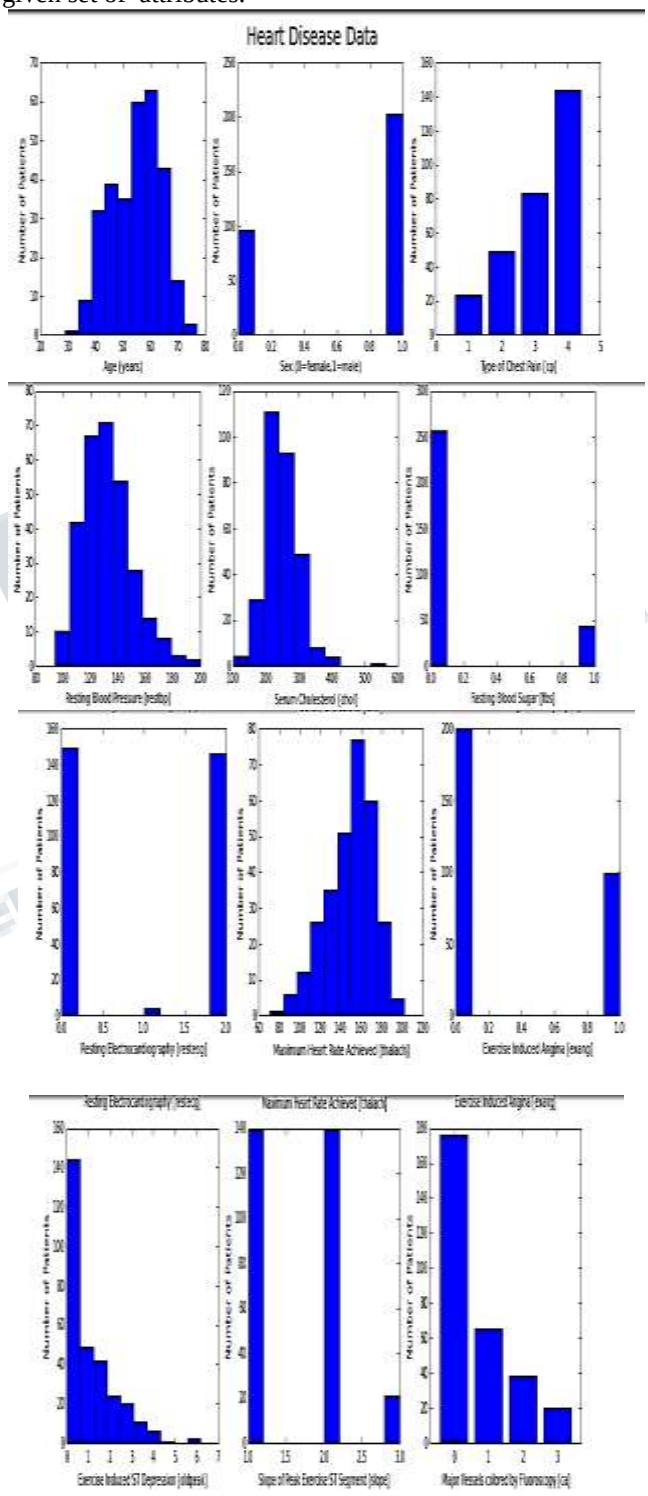
Accuracy of that classifier is calculated as follows

Accuracy = no. Of samples correctly classified in test data

Total no of samples in test data

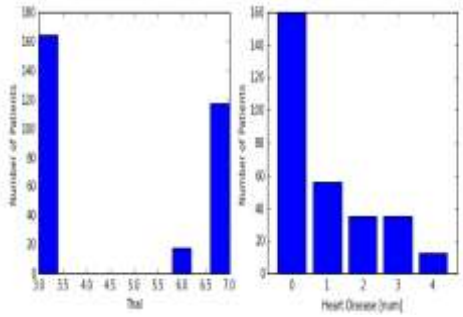
The performance this newly proposed algorithm is tested using Cleveland dataset which is downloaded from UCI repository.

Now the results are analyzed that by integrating both GA and KNN we can improve classification accuracy.Currently we tested it on Cleveland data set.The following graphs shows the analysis and prediction of heartdisease as the given set of attributes.

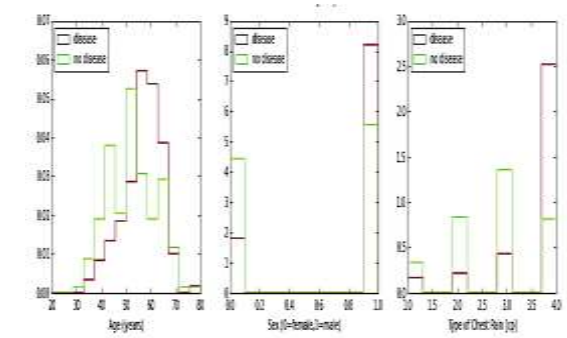


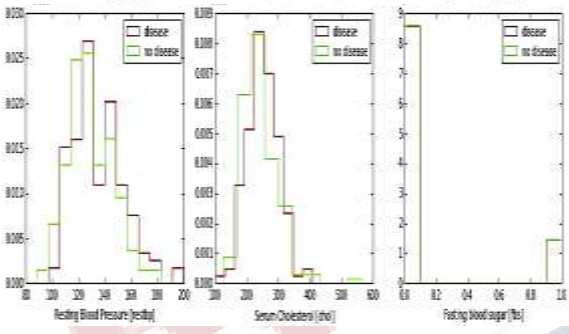
ISSN (Online) 2456 -1304

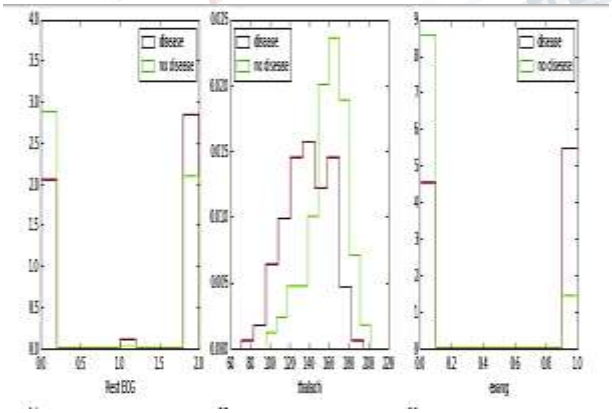
International Journal of Science, Engineering and Management (IJSEM)
Vol 3, Issue 4, April 2018

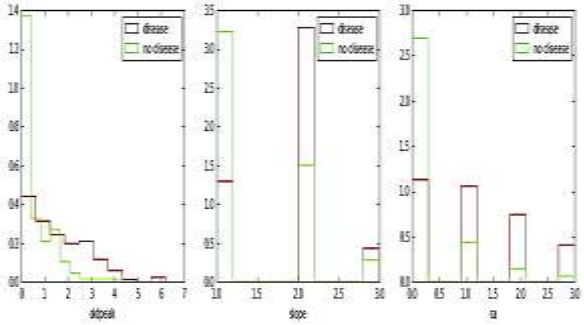


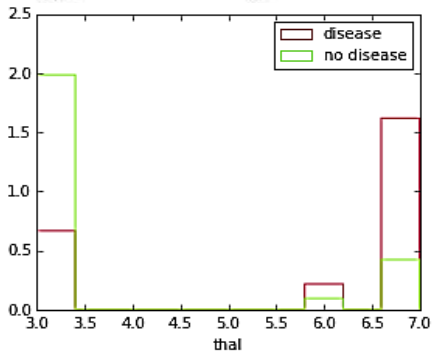
The following analysis shows that how many are diseased and not diseased based on their attributes.











II. CONCLUSION AND FUTURE WORK

In this paper we presented a hybrid algorithm for classification of heart disease. We have tested this by using Cleveland heart disease data set. This first implements GA for finding highly ranked attribute set and further classify it. This prediction model helps in a way to provide better diagnosis with only few attributes instead of large attribute set. Finally we can conclude that it can act as a powerful decision support system for predicting heart troubles for people.

REFERENCES

1. Classification of Heart Disease Using K- Nearest Neighbor and Genetic AlgorithmM.Akhil jabbar* B.L Deekshatulua Priti Chandra International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013.

2. Nikam S.S.A Comparative Study of Classification Tecniques in Data Mining Algorithms. Orient J.Comp. Sci. and Technol:8(1).

3. S.N Sivanandam,S.N Deepa,"Introduction to genetic algorithms"Springer(2008)

4. MA.Jabbar,B.L Deekshatulu,Priti chandra,"Heart disease prediction system using associative classification and genetic algorithm" pp183-192 Elsevier (2012)

5. MA.Jabbar,B.L Deekshatulu,Priti chandra."An evolutionary algorithm for heart disease prediction"CCIS,PP 378-389 , Springer(2012)

6. MA.Jabbar,B.L Deekshatulu,Priti chandra,"Prediction of Risk Score for Heart Disease using Associative classification and Hybrid Feature Subset Selection".In Conf ISDA,pp 628-634,IEEE(2013)

7. MA.Jabbar,B.L Deekshatulu,Priti chandra,"Knowledge discovery from mining association rules for heart disease prediction" pp45-53,vol 41,no 2 ,JATIT(2013)

8. WWW.ics.uci.edu/~mlearn

9. MA.Jabbar,B.L deekshatulu,priti chandra,"classification of heart disease using ANN and feature subset selection"GJCST, VOL 13,issue 3,version1.0 pp15-25(2013).

10. "An approach of Artificial Neural Networks for prediction of Generalized Anxiety Disorder" M.SriBala International Journal of Research in Computer Applications and Robotics"VOL.3,Issue 3,March 2015.

11. "Efficient Ensemble Classifiers for Prediction of Breast Cancer" M.SriBala,G.V.Rajya Lakshmi, International Journal of Advanced Research in Computer Applications and Robotics.Vol 6,Issue 3,March 2016.

12. E.Fix and J.Hodges,"Dicriminatory analysis ,non parametric discrimination:consistency properties".Technical report 4,USA,School of aviation medicine Randolph field texas(1951)

13. D.E Goldberg."Genetic algorithm in search .optimization and machine learning"Addison wesley(1989)

14. Dr saed sayad,"University of toronto http://chem-eng.utoronto.ca/~data mining

15. Nitin Bhatia ,vandana"Survey on nearest neighbor techniques"IJCSIS,Vol 80,no 2(2010)

16. M.SRI BALA, "An approach of artificial neural networks for prediction of generalized anxiety disorder" ,International Journal of Research in Computer Applications and Robotics.Vol3 Issue3 march 2015 pg 118-124.ISSN 2320-7345 impact factor 1.831.

17. M.Sri Bala ,G,V,Rajya Lakshmi " Efficient ensemble classifiers for prediction of breast cancer" for conference proceedings on Recent Trends in IT at PVPSIT Vijayawada IJARCSSE & special issue ISSN 2277-128X dated 24-03-2016.