International Journal of Technology Research and Management
ISSN (Online): 2348-9006
Vol 2 Issue 6 June 2015

Heart Attack Prediction Using Data Mining Classification Techniques: A Study

Amit Kisan Pagare[1], Vijay Kumar Verma[2]
Lord Krishna College of Technology M. Tech (CSE) 4th Sem. Indore M.P.[1]
Asst. Professor Dept CSE Lord Krishna College of Technology Indore M.P.[2]
amitpagare88@gmail.com[1]

Abstract: *- Classifications is one of the important data mining techniques. In we construct model to describe data classes. The derived model used training data set. New and better techniques are developed for data mining and to discover unknown pattern form historical data. Some of these Models are useful for medical practitioners to take better and successful decision. Heart attack is one of the important problems in medical science. Early predicting heart attack symptoms are an important issue. In this paper we represent comparative study over various data mining classification techniques like Neural Network Decision Tree, Bayesian Classification, and Classification by back propagation, Support Vector Machines (SVM), k-Nearest Neighbor and Genetic Algorithms (GA).*

Keyword: *- Prediction, Classification, Data Diagnosis, Heart Attack, Symptoms.*

1. Introduction

Constructing fast and accurate classifiers for large data sets is an important task in data mining and knowledge discovery. Classification predicts categorical class labels and classifies data based on the training set. Classification is two step processes[1,4].

1. Model construction

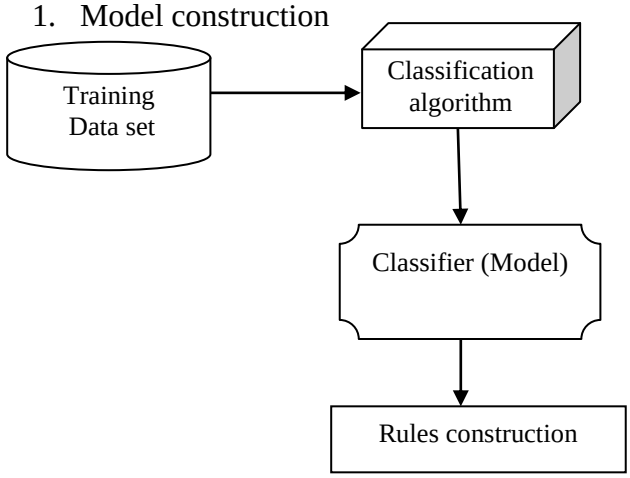

Figure 1 Model constructions

2. Model usage

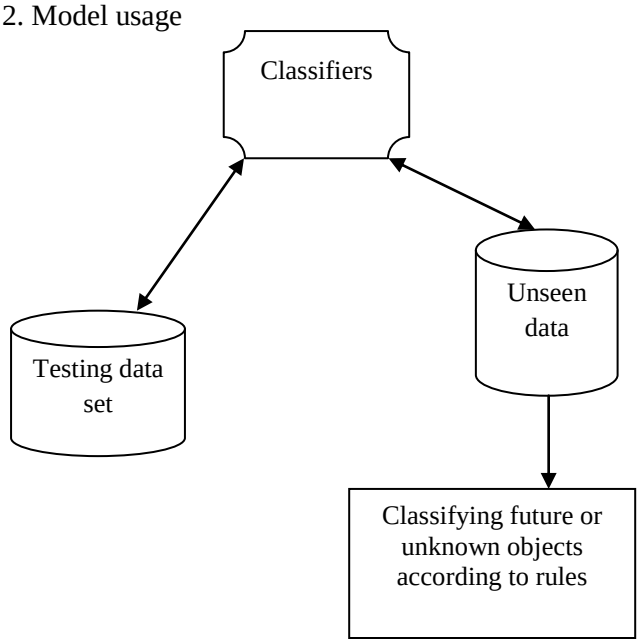

Figure 2 Model usage

2. Disease Classification Systems

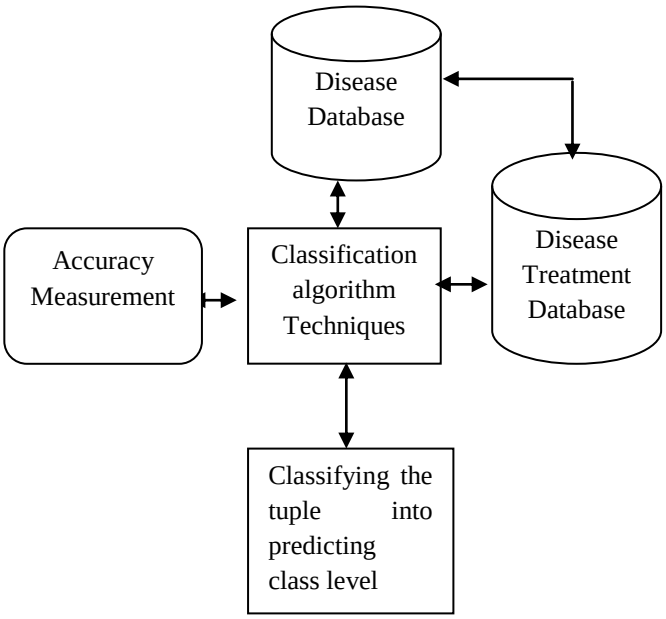

Figure 3 disease classification systems

Figure 3 show the disease classification system. This diagram contains disease database and disease treatment database for classification. A accuracy measurement is used in this diagram. Which hemp to classify a tuple accurately. There are several classification algorithms are exist we can use any one of them.

3. Common Classification Techniques

3.1 Decision Tree Classifier

Decision tree is general tree in which at the top node known as root of the tree. After the root next level exist intermediate nodes denotes a test for a particular branch. Leaf node denotes the class label which is a kind of yes or no decision. Using Decision Tree we select best path and move from root to leaf. Maximum information gain is selected for unique class separation.

Figure 4 simple diagrams for decision tree. At the top root of the tree denoted by A. B, C, D denotes internal nodes (non-leaf nodes) denote a test on a particular attribute and C1, C2, C3, C4, C5 denotes leaf node with a class label. To determine the class. Beginning with the root, successive internal nodes are visited until a leaf node is reached. At the root node and at each internal node, a test is applied[2,3].

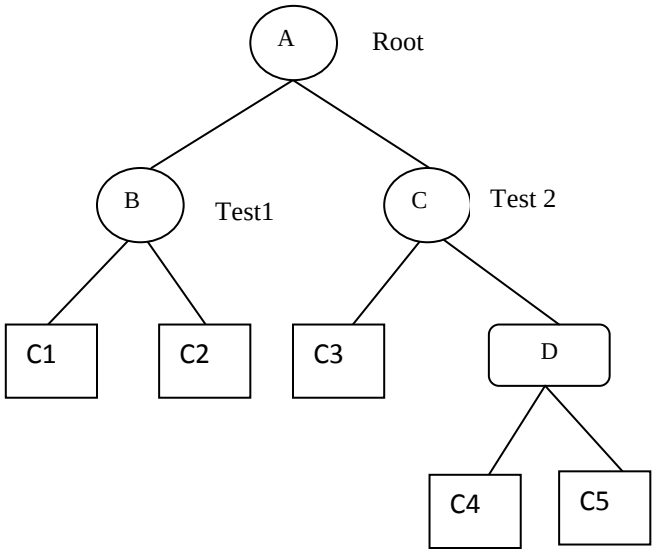

Figure 4 Decision tree

3.2 Naive Bayesian Classifiers

Simple Bayesian classifier is also known as Naive Bayesian classifier. Bayesian classifiers are kind of statistical classifiers which are based on Bayes' theorem. They are used to predict that a given tuple belongs to a particular class or not. This prediction is based on membership probabilities. Naïve Bayesian classifiers used class conditional independence means that the effect of an attribute value on a given class is independent of the values of the other attributes.Bayesian classifiers have also exhibited high accuracy and speed when applied to large databases[9,11].

3.3 Neural Network as a Classifier

An artificial neural network (ANN) is being applied to solve a number of real world problems. Neural networks

mine knowledgeable from data ware house. ANN is basically a trained tool which store and retrieve patterns to solve optimization problems. Pattern recognition and function Estimation abilities make ANN prevalent utility in data mining. Neural networks are well suited to problem like pattern recognition and forecasting[12,14].

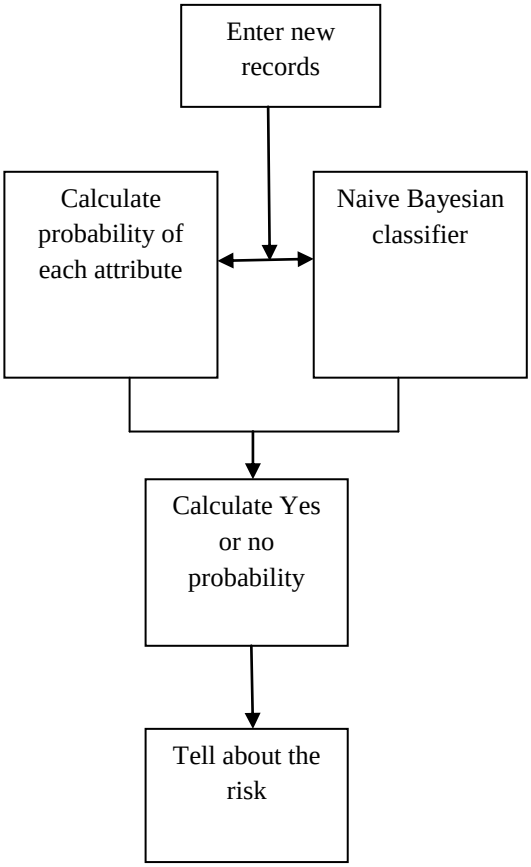

Figure 5 Bayes Classifier

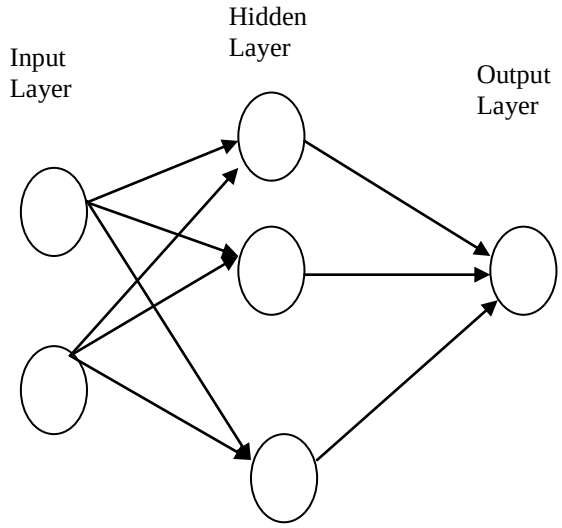

Figure 6 Single layers Neural Network

3.4 Support Vector Machine (SVM)

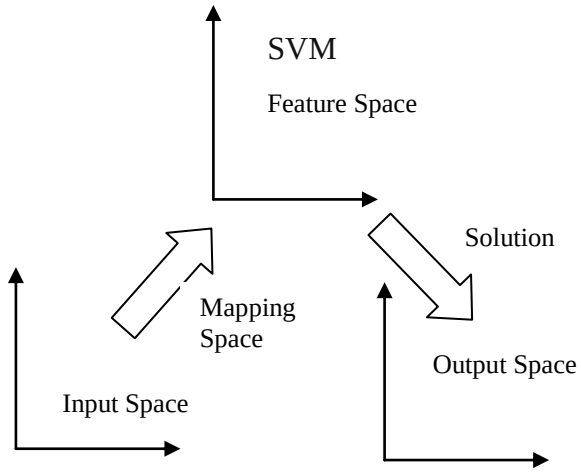

Figure 7 support vector machine classifier

SVM (Support Vector Machine) has been very effective method for classification and pattern recognition. It is considered a good classifier because of its high generalization performance without the need to add a priori knowledge. SVM first find the classification function to distinguish between members of the two

classes in the training data. For a linearly separable dataset, a linear classification function corresponds to a separating hyper plane is used that passes through the middle of the two classes two. One of the initial drawbacks of SVM is its computational inefficiency [9,15]. Figure 6 shows SVM .

4. Litrurere Review

In 2010 Sunita Soni O.P. Vyas proposed " associative Classifiers for Predictive Analysis in Health Care ". They proposed a approach using association rule mining and classification rule mining. The integration is done by focusing on mining a special subset of association rules known as CAR. By using association rule mining they construct classification. The proposed system can predict if the patient is likely to have a certain disease. They also discuss advanced associative classifiers being proposed[7] .

In 2010 N. Suneetha, Ch.V.M.K.Hari, V.Sunil Kumar proposed "Modified Gini Index Classification". They modified decision tree method to overcome the known problems of Gini-based. They normalize the Gini indexes by taking information about the splitting status of all attributes. Instead of using the Gini index for attribute selection ratios of Gini indexes are used and their splitting values in order to reduce the biases[5].

In 2011 Mai Shouman, Tim Turner, proposed "Using Decision Tree for Diagnosing Heart Disease Patients". In this paper they represent that Bagging algorithm and decision tree are two efficient algorithms in the diagnosis of heart disease patients. Decision Tree is one of the successful data mining techniques used in the diagnosis of heart disease. Yet its accuracy is not perfect. They tested combinations of discretization, decision tree type and voting to identify a more robust, more accurate method[8].

In 2011 Jyoti Soni, Uzma Ansari, proposed "Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers". In Weighted Associative Classifier (WAC), different weights are assigned to different attributes according to their predicting capability. It has already been proved that the Associative Classifiers are performing well than traditional classifiers approaches such as decision tree and rule induction[6].

In 2012 Sunita Soni And O.P.Vyas proposed "Fuzzy Weighted Associative Classifier: A Predictive Technique For Health Care Data Mining". They proposed a new Fuzzy Weighted Associative Classifier (FWAC). The naive approach can be used to generating strong rules instead of weak irrelevant rules. They proposed a theoretical model to introduce new classifier. They proposed a new foundational approach to Fuzzy Weighted Associative Classifiers where quantitative attributes are discritized to get transformed binary database[10].

In 2014 N S Nithya, and K Duraiswamy proposed "Gain ratio based fuzzy weighted association rule mining classifier for medical diagnostic interface". They enhanced approach, called gain ratio based fuzzy weighted association rule mining, is thus proposed for distinct diseases and also increase the learning time of the previous one[13].

5. Benefits and Limitations

Methods	Benefits	Limitations
Decision Tree	1. Needs domain facts in the construction of decision tree. 2. Reduces the	1. Restricted only output attribute. 2. Generate categorical output. 3. Performance of

	ambiguity of decisions .. 3. high dimension. Data can easily	classifier is depend upon the type of dataset.
Bayesian Belief Network	1.Makes computations easier. Speed and accuracy are better for large datasets	1. When dependency exists between variables then it does not give accurate results.
Neural Network	1. Easily identify relationships among dependent and independent variables. 2. Noisy data can also be handled.	1. Over-fitting. 2. ANN network is difficult to interpret and require high processing time.
Support Vector Machine	1. Provide better accuracy. 2. Easily handle nonlinear data points.	1. Computationally expensive. 2. takes more time As compare to other methods

6. Conclusion

The performance of classification techniques depends on the type of dataset. Classification techniques give benefit to doctor, healthcare insurers, patients and organizations who are engaged in healthcare industry. These techniques are compared on basis of Sensitivity, Specificity, Accuracy, Error Rate, True Positive Rate and False Positive Rate. The objective of each technique is to predict more accurately the presence of heart disease with reduced number of attributes.

7. Reference

[1] Divya Tomar and Sonali Agarwal " A survey on Data Mining approaches for Healthcare" International Journal of Bio-Science and Bio-Technology Vol.5, No.5 (2013), pp. 241-266.

[2] Bangaru Veera Balaji and Vedula Venkateswara Rao "Improved Classification Based Association Rule Mining" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 5, May 2013

[3] V. Krishnaiah and Dr. G. Narsimha proposed " Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques " (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (1) 2013, 39 – 45

[4] Shamsher Bahadur Patel, Pramod Kumar Yadav and Dr. D. P.Shukla " Predict the Diagnosis of Heart Disease Patients Using Classification Mining Techniques" IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS) e-ISSN: 2319-2380, p-ISSN: 2319-2372. Volume 4, Issue 2 (Jul. - Aug. 2013),

[5] N. Suneetha ,Ch.V.M.K.Hari and Sunil Kumar " Modified Gini Index Classification: A Case Study Of Heart Disease Dataset" (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 06, 2010, 1959-1965

[6] Jyoti Soni, Uzma Ansari, Dipesh Sharma and Sunita Soni "Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers**"** Jyoti Soni et al. / International Journal on Computer Science and Engineering (IJCSE) ISSN : 0975-3397 Vol. 3 No. 6 June 2011

[7] Sunita Soni and O.P.Vyas Using Associative Classifiers for Predictive Analysis in Health Care Data Mining "International Journal of Computer Applications (0975 – 8887) Volume 4 – No.5, July 2010

[8] Mai Shouman, Tim Turner, Rob Stocker "Using Decision Tree for Diagnosing Heart Disease Patients" Proceedings of the 9-th Australasian Data Mining Conference (AusDM'11), Ballarat, Australia

[9] M. Akhil jabbar, Dr B.L Deekshatulu and Dr Priti Chandra " Heart Disease Classification Using Nearest

Neighbor Classifier With Feature Subset Selection" Computer Science and Telecommunications 2013|No.3(39)

[10] Sunita Soni and O.P.Vyas "Fuzzy Weighted Associative Classifier: A Predictive Technique For Health Care Data Mining" International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.1, February 2012

[11] Chaitrali S. Dangare and Sulabha S. Apte, PhD. "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques" International Journal of Computer Applications (0975 – 888) Volume 47– No.10, June 2012

[12] M. Akhil Jabbar, B.L Deekshatulu & Priti Chandra "Classification of Heart Disease using Artificial Neural Network and Feature Subset Selection" Global Journal of Computer Science and Technology Neural & Artificial Intelligence Volume 13 Issue 3 Version 1.0 Year 2013 Type: Double Blind Peer Reviewed International Research Journal Publisher: Global Journals Inc. (USA) Online ISSN: 0975-4172 & Print ISSN: 0975-4350

[13] N S Nithya and K Duraiswamy Gain ratio based fuzzy weighted association rule mining classifier for medical diagnostic interface Vol. 39, Part 1, February 2014, pp. 39–52. Indian Academy of Sciences

[14] M.Akhil jabbar, Dr.Priti Chandrab , Dr.B.L Deekshatuluc "Heart Disease Prediction System using Associative Classification and Genetic Algorithm" International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies-ICECIT, 2012

[15] A. Anushya and A. Pethalakshmi "A Comparative Study of Fuzzy Classifiers With Genetic On Heart Data" International Conference on Advancement in Engineering Studies & Technology, ISBN : 978-93-81693-72-8, 15th JULY, 2012, Puducherry