IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

A Survey on Heart Disease Forecasting using Hybrid Technique in Data Mining

Vivek Barot[1], Prof. Jay Vala[2]

PG Scholar, Department of IT, G.H. Patel College of Engineering and Technology, V.V. Nagar, Anand[1]

Assistant Professor, Department of IT, G.H. Patel College of Engineering and Technology, V.V. Nagar, Anand[2]

Abstract: Data mining is a process of extracting information and discovering useful patterns from the vast amount of data. It is also known as Knowledge Discovery from Data (KDD).The healthcare industry gathers very large amounts of healthcare data which leads to the need for powerful data analysis tools to extract useful information. Disease diagnosis is one of the applications where data mining techniques are showing successful results. Researchers are working on several statistical analysis and data mining techniques to enhance the disease diagnosis accuracy in medical healthcare. Heart disease diagnosis is evaluated as the complex tasks in the medical field. Every year the number of death is increasing because of heart disease. Different data mining techniques individually give low accuracy for disease diagnosis. Researchers are examining the effect of combining more than one technique showing enhanced results in the diagnosis of heart disease. In this paper, we survey different papers in which single techniques or combination of different data mining techniques are used to the forecasting of heart disease so that we can recognize the technique with high accuracy for future research.

Keywords: Data mining; Heart disease forecasting; Data mining techniques.

I. INTRODUCTION

The main goal of Data mining is to find interesting patterns and knowledge from huge amount of data. Data mining can be known by various names like knowledge discovery from databases (KDD), knowledge extraction, data or pattern analysis, data dredging etc.[1] There are two strategies in Data mining : supervised and unsupervised learning. Supervised learning uses a training set to learn model variables while Unsupervised learning does not use any training set[15]. Supervised learning is also called as Classification. The aim of classification is to accurately predict the target class label of an object for which the class label is unknown. There are two steps in classification first is model construction in which a set of predetermined classes from database generates training set and second is Model usage where training set is used to predict label of unknown objects. There are many application of classification such as customer segmentation, business modeling, marketing, credit analysis, and healthcare and weather forecasting. Biomedical data is used to classify or forecast many diseases such as diabetes, stroke, cancer, and heart disease [8].

In human body, heart is the main part/organ which is made up of muscles and nerves. Any failure or defect in the heart may result in sudden death. In the USA statistics has shown that about 800,000 people die of heart disease every year [7]. In 2003 approximately 17.3 million people died all over the world and out of this, 10 million were only due to the CHD (coronary heart disease). There are many kind of classification techniques such as K-nearest neighbour, Decision Trees like CART, C4.5, J48, ID3 algorithm ,SVM, Artificial Neural network, Naive Bayes etc. Which are used to forecast the heart diseases but gives low accuracy individually so we need the help of boosting and bagging techniques or combination of different techniques to improve their performance [9].

The remaining part of this paper is organized as follows: Section II Literature Survey about the recent contribution made in heart diseases forecasting, Section III is about Performance analysis of different classifiers, Section IV is about Conclusion and future scope.

II. LITERATURE SURVEY

There are many approaches have been used to forecast heart diseases using data mining techniques. Some of this is presented here.

Bashir S et al., [5] proposed An Ensemble based Decision Support Framework for Intelligent Heart Disease Diagnosis using majority vote based technique which uses Naïve Bayes; Decision tree and Support vector machine classifiers. Proposed framework gives 81.82% classification accuracy.

Olaniyi EO et al., [7] developed an intelligent system to effectively diagnosis patient to avoid misdiagnosis. This intelligent system is a model on multilayer neural network trained with back propagation and simulated on feed forward neural network. The authors used the heart disease dataset of UCI statlog heart disease. Compare to other algorithms like KNN, Decision Tree and Naïve Bayes this model obtained 85% recognition rate which is a lot better.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

Shouman M et al.,[8] recognize gaps in the research on cardiovascular disease diagnosis and forecasting and proposed a model to consistently shut those gaps to find if applying data mining techniques to cardiovascular disease treatment knowledge will give as reliable performance as that achieved in identification cardiovascular disease patients. Fida B et al.,[10] proposed framework of SVM classifier ensemble and results are optimized using Genetic Algorithm technique to get better classification accuracy compared to other classifiers. Here they used four different datasets, three of them (Cleveland, Statlog, SPECT) are obtained from UCI machine learning repository and the fourth one is taken from S.african dataset.The maximum accuracy obtained by their proposed method is 98.63%. Amma NB.[11] proposed a system for forecasting the risk of heart disease. The system is a combination of neural network and genetic algorithm. By using genetic algorithm weights of the neural network are determined. Proposed system obtained 94.17% classification accuracy. Anushya A et al.,[12] examined the performance of four classifiers named Decision Trees, K-means, Naive Bayes and neural network with the combination of Fuzzy Logic on heart data to improve the accuracy of the classifier. The dataset for experiments are taken from UCI machine learning repository and it is implemented in MATLAB. The Fuzzy K-means technique gives an accuracy of 99.2650% which is batter then other three classifiers. Chen, A.H et al. [13] developed a Heart disease prediction system (HDPS) for distinguishing cardiovascular disease in patients more accurately. A prophetical model is employed to diagnose the heart disease.Statistics and machine learning are two main techniques that are employed in the proposed system. The algorithmic program has three steps: the first step is feature selection, the second step is artificial neural network model for classification and the last one is user-friendly Heart disease prediction system. Learning Vector quantisation (LVQ) which is supervised classification algorithm is then applied on a dataset for training. To check the accuracy of results Receiver Operating Characteristic curve(ROC) is used. The prediction accuracy of heart disease prediction system is near 80%

Peter TJ et al.,[14] examined the performance of various classification techniques such as Naive Bayes, Decision Tree, K-NN and Neural Networks. The dataset is reduced by CFS attribute selection method and applied on different classifiers, naive Bayes gives better classification accuracy for heart disease forecasting.

Rajkumar A et al.,[16] examined the performance of Naive Bayes, K-NN, Decision tree classifiers. Classification-id done by using Tanagra tool and data is evaluated using 10-fold cross validation technique. In last the result of Naive Bayes, K-NN, Decision tree classifiers are compared. Naive Bayes classifier gives better accuracy then other two classifiers.

Anbarasi M et al.,[19] proposed system which to reduce the size of dataset Genetic algorithm is used to select the attributes. By using CFS with Genetic algorithm 6 attributes are selected from 13 attributes, then this reduced dataset is applied on Naive Bayes, Classification by clustering and Decision Tree classifiers. Decision Tree classifier gives better accuracy compare to other two classifiers.

My Chau Tu et al.,[20] proposed a system for identifying the coronary artery disease of a patient by using a decision tree C4.5 algorithm, bagging with decision tree C4.5 algorithm and bagging with Naïve Bayes algorithm. To compute confusion matrix of each model and to evaluate the performance they used 10-fold cross validation technique. Bagging algorithms gives a better result than other algorithms especially the bagging with Naïve Bayes gives the best result and precision accuracy up to 82.50%.

My Chau Tu et al.,[21] examined the performance of the bagging algorithm and compare it with the decision tree algorithm. The dataset of coronary artery disease (CAD) is taken from UCI machine learning repository. The bagging algorithm gives better performance and accuracy than the decision tree on coronary artery disease dataset.

Kangwanariyakul Y et al.,[22] developed classification models for identifying Ischemic Heart Disease(IHD) patients using the Bayesian neural network(BNN),Back-propagation neural network(BPNN), the probabilistic neural network(PNN) and the support vector machine(SVM). Magneto cardiogram(MCG) is used to detect electro-physiological activity of the myocardium. Compare to other classification techniques BPNN and BNN gives the highest classification accuracy of 78.43 %.

III. PERFORMANCE ANALYSIS OF DIFFERENT CLASSIFIERS

Different data mining techniques have been used to help health care professionals in the diagnosis of heart disease. Table 1 shows a various data mining techniques used in the diagnosis of heart disease over different heart disease datasets [8].

TABLE I DATA MINING TECHNIQUES USED ON DIFFERENT HEART DISEASE DATASETS

Author	Year	Technique	Accuracy
Yan, et al.	2003	Multilayer Perceptron	63.6%
Andreeva, P.	2006	Naïve Bayes	78.563%
		Decision Tree	75.738%
		Neural network	82.773%
		Kernel density	84.444%

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

Palaniappan, et al.	2007	Naïve Bayes	95%
		Decision Tree	94.93%
		Neural network	93.54%
De Beule, et al.	2007	Artificial neural network	82%
Tantimong colwata, et al.	2007	Automatically Defined Groups	67.8%
		Immune Multi-agent Neural Network	82.3%
Sitar-Taut, et al.	2009	Naïve Bayes	62.03%
		Decision Tree	60.40%
Tu, et al.	2009	J4.8 Decision Tree	78.91%
		Bagging algorithm	81.41%
Rajkumar, et al.	2010	Naïve Bayes	52.33%
		KNN	45.67%
		Decision list	52%
Srinivas, et al.	2010	Naïve Bayes	84.14%
		One Dependency Augmented Naïve Bayes classifier	80.46%
Kangwanariyakul, et al.	2010	Back-propagation neural network	78.43%
		Bayesian neural network	78.43%
		probabilistic neural network	70.59%
		linear support vector machine	74.51%
		polynomial support vector machine	70.59%
		radial basis function kernel support vector machine	60.78%
Anbarasi, et al.	2010	Genetic with Naïve Bayes	99.2%
		Genetic with Decision Tree	96.5%
		Genetic with Classification via clustering	88.3%
Fida B et al.	2011	SVM-L	93.12%
		SVM-P	89.65%
		SVM-RBF	89.65%
Anushya A et al.	2011	Fuzzy Decision Tree	91.2530%
		Fuzzy NaiveBayes	93.3040%
		Fuzzy neural network	95.0830%
		Fuzzy Kmeans	99.2650%
Peter TJ et al.	2012	NaIve Bayes	83.70%
		Decision Tree	76.66%
		K-NN	75.18%
		Neural network	78.148%
Bashir S et al.	2014	NaIve Bayes	78.79%
		Decision Tree	72.73%
		SVM	75.76%
Olaniyi EO et al.	2015	K-NN	45.67%
		Decision Tree	84.35%
		NaIve Bayes	82.31%
		BPNN	85%

IV. CONCLUSION AND FUTURE SCOPE

In this survey paper, we have studied different data mining techniques for forecasting heart disease. From this survey, we got the information about how to apply various data mining technique to forecast the heart disease. Earlier exiting system was designed with a single algorithm which is not providing sufficient accuracy, nowadays researchers work on hybrid technique to acquire more accuracy. Applying hybrid data mining techniques has shown good results in the forecasting of heart disease, so in future an Intelligent system can be developed by using hybrid data mining techniques to get better accuracy.

ACKNOWLEDGMENT

Vivek Barot remains thankful to **Prof. Jay Vala** (Department Of Information Technology, GCET) for their useful discussions & suggestions during the preparation of this technical paper.

REFERENCES

[1] Han, Jiawei, Jian Pei, and Micheline Kamber. Data mining: concepts and techniques. Elsevier, 2011.
[2] Lakshmi BN, Raghunandhan GH. A conceptual overview of data mining. InInnovations in Emerging Technology (NCOIET), 2011 National Conference on 2011 Feb 17 (pp. 27-32). IEEE.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

[3] Kesavaraj G, Sukumaran S. A study on classification techniques in data mining. InComputing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on 2013 Jul 4 (pp. 1-7). IEEE.
[4] Gandhi M, Singh SN. Predictions in heart disease using techniques of data mining. InFuturistic Trends on Computational Analysis and Knowledge Management (ABLAZE), 2015 International Conference on 2015 Feb 25 (pp. 520-525). IEEE.
[5] Bashir S, Qamar U, Javed MY. An ensemble based decision support framework for intelligent heart disease diagnosis. InInformation Society (i-Society), 2014 International Conference on 2014 Nov 10 (pp. 259-264). IEEE.
[6] Muhammed LA. Using data mining technique to diagnosis heart disease. InStatistics in Science, Business, and Engineering (ICSSBE), 2012 International Conference on 2012 Sep 10 (pp. 1-3). IEEE.
[7] Olaniyi EO, Oyedotun OK, Helwan A, Adnan K. Neural network diagnosis of heart disease. In2015 International Conference on Advances in Biomedical Engineering (ICABME) 2015 Sep 16 (pp. 21-24). IEEE.
[8] Shouman M, Turner T, Stocker R. Using data mining techniques in heart disease diagnosis and treatment. InElectronics, Communications and Computers (JEC-ECC), 2012 Japan-Egypt Conference on 2012 Mar 6 (pp. 173-177). IEEE.
[9] Dewan A, Sharma M. Prediction of heart disease using a hybrid technique in data mining classification. InComputing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on 2015 Mar 11 (pp. 704-706). IEEE.
[10] Fida B, Nazir M, Naveed N, Akram S. Heart disease classification ensemble optimization using genetic algorithm. InMultitopic Conference (INMIC), 2011 IEEE 14th International 2011 Dec 22 (pp. 19-24). Ieee.
[11] Amma NB. Cardiovascular disease prediction system using genetic algorithm and neural network. In2012 International Conference on Computing, Communication and Applications 2012 Feb 22 (pp. 1-5). IEEE.
[12] Anushya A, Pethalakshmi A. A comparative study of fuzzy classifiers on heart data. In3rd International Conference on Trendz in Information Sciences & Computing (TISC2011) 2011 Dec 8 (pp. 17-21). IEEE.
[13] Chen AH, Huang SY, Hong PS, Cheng CH, Lin EJ. HDPS: heart disease prediction system. In2011 Computing in Cardiology 2011 Sep 18 (pp. 557-560). IEEE.
[14] Peter TJ, Somasundaram K. An empirical study on prediction of heart disease using classification data mining techniques. InAdvances in Engineering, Science and Management (ICAESM), 2012 International Conference on 2012 Mar 30 (pp. 514-518). IEEE.
[15] Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. In2008 IEEE/ACS International Conference on Computer Systems and Applications 2008 Mar 31 (pp. 108-115). IEEE.
[16] Rajkumar A, Reena GS. Diagnosis of heart disease using datamining algorithm. Global journal of computer science and technology. 2010 Dec;10(10):38-43.
[17] Bouali H, Akaichi J. Comparative Study of Different Classification Techniques: Heart Disease Use Case. InMachine Learning and Applications (ICMLA), 2014 13th International Conference on 2014 Dec 3 (pp. 482-486). IEEE.
[18] Srinivas K, Rani BK, Govrdhan A. Applications of data mining techniques in healthcare and prediction of heart attacks. International Journal on Computer Science and Engineering (IJCSE). 2010 Feb;2(02):250-5.
[19] Anbarasi M, Anupriya E, Iyengar NC. Enhanced prediction of heart disease with feature subset selection using genetic algorithm. International Journal of Engineering Science and Technology. 2010 Jan 1;2(10):5370-6.
[20] Tu MC, Shin D, Shin D. A comparative study of medical data classification methods based on decision tree and bagging algorithms. InDependable, Autonomic and Secure Computing, 2009. DASC'09. Eighth IEEE International Conference on 2009 Dec 12 (pp. 183-187). IEEE.
[21] Tu MC, Shin D, Shin D. Effective diagnosis of heart disease through bagging approach. In2009 2nd International Conference on Biomedical Engineering and Informatics 2009 Oct 17 (pp. 1-4). IEEE.
[22] Kangwanariyakul Y, Naenna T, Nantasenamat C, Tantimongcolwat T. Data mining of magnetocardiograms for prediction of ischemic heart disease.EXCLI Journal July 30, 2010 9:82-95