



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

A Literature Survey on Mining Health Data to Predict Heart Disease

P. Priyanga¹, Aparna N Venu², Latha Eshappa², M Archana², Navya Shree D P²

Assistant Professor, Dept. of CSE, KSIT, Bengaluru, Karnataka, India¹

B.E Students, Dept. of CSE, KSIT, Bengaluru, Karnataka, India²

ABSTRACT: Almost 19% of deaths in India in the year 2015 were attributed to Cardiovascular Diseases (CVD). Early detection and prediction of CVD is very important for patients' treatment and doctors' diagnose which can help to reduce mortality. Computational intelligence and Data mining plays an important role in the field of heart disease prediction. Thus, there arises a need to develop a support system for detecting heart diseases in a patient. Unexpected acute events have resulted in much affliction as well as high treatment costs. The latter are now reaching unsustainable levels and are becoming huge burdens even for developed countries. Early prediction and intervention would therefore be of huge benefit to society. In this paper, we propose efficient data mining techniques to select the best features with the lowest costs and shortest times and machine learning algorithms to achieve the accuracy. This technique will help to reduce the work load and cost for patients as well as health care unit.

KEYWORDS: Cardiovascular disease, Computational Intelligence, Data Mining, Machine Learning algorithms.

I. INTRODUCTION

Heart disease is the form of cardio-vascular disease which is prevailing in the whole world like a wave and becoming a bigger cause of deaths. Latest statistics suggest that in India, there are roughly 30 million heart patients and according to Mr. KK Aggarwal, the President of Heart Care Foundation of India this number will continue to increase due to things like stress, unhealthy eating habits, lack of physical exercise, lack of sleep and dependence on alcohol and cigarettes [14].

Cardiovascular diseases (CVD) are a class of diseases that involve heart and [blood vessels](#). Cardiovascular diseases include [coronary artery diseases](#) (CAD) such as [angina](#) and [myocardial infarction](#) (commonly known as a heart attack) [12]. Earlier methods of predicting Heart disease included doctor's examination or number of medical tests such as ECG, Stress Test, and Heart MRI etc. Nowadays, Health care industries contain large amount of health care data, which contains hidden information. This information is useful for making effective decisions.

Nowadays, Data mining techniques are playing a vital role in the biomedical field as they are used to explore, analyze and extract medical data using complex algorithms in order to discover unknown patterns [13]. Data Mining has two primary goals- prediction and description. Prediction involves using some variables or fields in the data set to predict unknown or future values of other variables of interest. Description, on the other hand, focuses on finding patterns describing the data that can be interpreted by humans.

Researchers are using data mining techniques for the diagnosis of many diseases such as heart disease, diabetes, stroke and cancer and many data mining techniques together with machine learning algorithms have been used in the diagnosis of heart disease with good accuracy.

Machine learning algorithms are useful techniques to extract flexible and comprehensible knowledge from huge datasets. These are less complicated to implement and their results are more easily understood to users. Machine learning techniques deal with a mix of quantitative, qualitative, missing or noisy data so common on engineering.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirce.com

Vol. 5, Issue 5, May 2017

II. RELATED WORK

There have been a huge number of articles conducted on Data sets of cardiovascular diseases based on data-mining methods.

Ankita Dewan and Meghna Sharma proposed a prototype which can determine and extract unknown knowledge (patterns and relations) related with heart disease from a past heart disease database record. It can solve complicated queries for detecting heart disease and thus assist medical practitioners to make smart clinical decisions which traditional decision support systems were not able to. By providing efficient treatments, it can help to reduce costs of treatment. This work could use an efficient optimizing technique to further improve its accuracy and apply in the predictions of various applications [4].

Majid Ghonji Feshki and Omid Sojoodi Shijani of Islamic Azad University used feature ranking on effective factors of disease related to Cleveland clinic database and by using Particle Swarm Optimization as well as Neural Network Feed Forward Back Propagation, 13 effective factors reduced to 8 optimized features in terms of cost and accuracy. The assessment of selected features of classified methods also showed that PSO method along with Neural Networks of Feed Forward Back-Propagation has the best accurate criteria of the rate of 91.94% on these features. This method can be extended to increase the accuracy using Data Mining techniques. Theresa Princy. R and J. Thomas of Christ University faculty of engineering gave the survey about different classification techniques used for predicting the risk level of each person based on age, gender, Blood pressure, cholesterol, pulse rate. The patient risk level is classified using data mining classification techniques such as Naïve Bayes, KNN, Decision Tree Algorithm, and Neural Network. etc., Accuracy of the risk level is high when using more number of attributes. In future, the numbers of attributes could be reduced and accuracy would be increased using some other algorithms [5].

Ahmad Mohawish, Ragini Rathi, and Vibhanshu Abhishek explored the application of the Framingham Risk Model to predict CHD (Coronary Heart Disease) risk using a limited set of attributes present in a health risk assessment (HRA) dataset from a digital health company. HRAs often fail to capture all the needed attributes of the Framingham Model, that significantly affect CHD risk [6]. In order to improve the accuracy of prediction while continuing to achieve high HRA completion rate, it is recommended that the factors used in Framingham study should be chosen in an accurate way.

Raid Lafta, Yan Li and Vincent S. Tseng of University of Southern Queensland developed an intelligent recommender system, which uses an innovative time series prediction algorithm to provide recommendations to heart disease patients in the tele-health environment. Based on analytics of each patient's medical tests in records, the system provides the patient with decision support for necessity of medical tests [7]. It can further be improved to predict the ability of the algorithm in order to enhance the precision and workload saving performance.

M.A. Jabbar, B.L. Deekshatulu and Priti Chandra attempted to increase the predictive accuracy of the Naïve Bayes to classify heart disease data by using discretization method and genetic search. Genetic search is used for optimization problem. The proposed approach can further be enhanced [8].

Syed Umar Amin, Kavita Agarwal, Dr. Rizwan Beg of Integral University presents a technique for prediction of heart disease using major risk factors. This technique involves two most successful data mining tools, neural networks and genetic algorithms. The hybrid system implemented uses the global optimization advantage of genetic algorithm for initialization of neural network weights. The learning is fast, more stable and accurate as compared to backpropagation. The system was implemented in Matlab and predicts the risk of heart disease with an accuracy of 89% which can further be improved using other techniques [9].

B.V. Baiju and R.J. Remy Janet of Hindustan University, have applied Naïve Bayes data mining classifier technique which produces an optimal prediction model using minimum training set. Data mining is the analysis step of the Knowledge Discovery in Databases process (KDD). Data mining involves use of techniques to find underlying structures and relationships in a large database. Using medical profile such as age, sex, blood pressure and blood sugar it can easily be predicted about the likelihood of patients getting heart disease [4]. The paper has evaluated the performance of new classification approach that uses the experienced Doctor's knowledge to assign the weight to each attribute. More weight is assigned to the attribute having high impact on disease prediction. The accuracy can be improved by adding other required features [10].

J.E. Nalavade, M.L. Gavali, N.D. Gohil and S.C. Jamale of Sinhgad Institute of Technology uses neural network, a professional methodology for the extraction of easiest patterns from the heart disease warehouses for heart attack

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 5, May 2017

prediction. Data warehouse is preprocessed in sequence to make it easy for the mining process. Processing gets finished, then heart disease warehouse is clustered with aid of K-means clustering algorithm, which will extract data, appropriate to heart attack from warehouse [11]. The frequent patterns applicable to heart disease are mined with aid of the algorithm from data extracted. The patterns important to heart attack prediction are selected on basis of the significant weight. The neural network is well trained with selected significant patterns for effective heart attack prediction system. It has implemented the Multilayer Neural Network with Back-propagation training algorithm. Results obtained have illustrated that designed prediction system is capable of predicting the heart attack more effectively. The future work is intended to apply on high performance client server or parallel processing computer and optical neural network as a classifier model [3].

III. PROPOSED FRAMEWORK

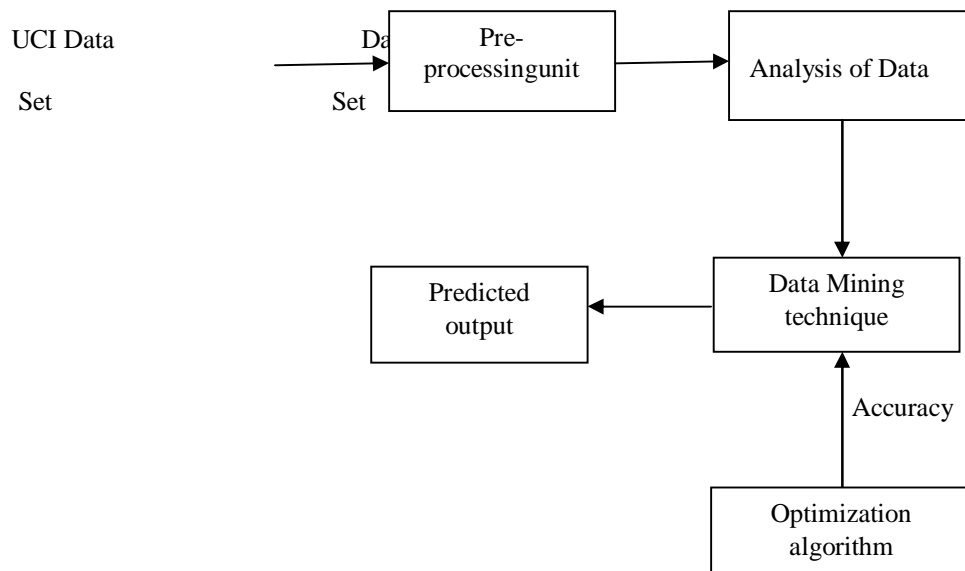


Fig 1: Proposed Framework for prediction of heart disease.

There have been numerous ways to predict the risk of heart disease, Fig 1 shows the basic flow of prediction that is followed.

A. Database of Patients:

The database of patients is collected and the most accurate and major attributes and feature which help in predicting heart disease such as: age(in years),sex(m/f),chest pain type(typical angina, atypical angina, non angina, asymptomatic),resting blood pressure(in mmHg),cholesterol(mg/dl), high fasting blood sugar(1/0),resting electrocardiographic results(1/0), maximum heart rate achieved, resting heart rate, exercise induced angina(1/0), St depression induced by exercise relative to rest,number of color vessels, obesity, thal(3=normal, 6= fixed defect, 7=reversible defect) are considered.

B. Analysis of Data:

It is one of the most important steps as the data in the database contain most of the redundant and noisy data. So by analysis of data, data cleaning, data integration can be performed to fill up missing values, remove the redundant data because handling missing value and redundant data would lead to incorrect output.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 5, May 2017

C. Optimization of Algorithm:

Various algorithms can be applied here to explore the best attribute which will evaluate the fitness value which is assigned to each attribute(individual). Algorithms which can be used can be PSO, Genetic Algorithm, Ant Colony Optimization, and Artificial Neural Network or may be the hybrid of any of these algorithms.

D.Prediction Engine:

Predicts whether the person has a heart disease or will suffer in future.

IV. CONCLUSION

Diagnosis of heart disease by using machine learning methods is one of the challenges in the health field. Various Data mining techniques and classifiers are discussed in many studies which are used for efficient heart disease prediction. Different technologies give different precision depending on a number of attributes considered. These systems improve the quality of clinical evidence-based decisions and help reduce the financial and timing cost taken by patients. All the techniques discussed could prove valuable in helping to address some of the challenges associated with reducing the expected healthcare spending due to CVD by informing, engaging and empowering individuals to actively participate in modifying their most significant risk factors for CVD. Hence, understanding the usefulness of data mining for early diagnosis of various heart diseases becomes important.

V.ACKNOWLEDGMENT

"Authors would like to thank VGST (Vision Group on Science and Technology), Government of Karnataka, India for providing infrastructure facilities through the K-FIST Level II project at KSIT, CSE R&D Department, Bengaluru".

REFERENCES

1. M.AkhiJabbar, B.L. Deekshatulu and Priti Chandra, "Classification of Heart Disease using Artificial Neural Network and Feature Subset selection", Global Journal of Computer Science and Technology Neural and Artificial Intelligence, Volume 13, issue 3 version 1.0 year 2013.
2. K. Sudhakar, Dr. Manimekalai, "Study of Heart disease prediction using Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, issue 1, pp.1157-60, January 2014.
3. B.V >Baijuand R.J. Remy Janet, "A survey on Heart Disease Diagnosis and Prediction using Naïve Bayes in Data Mining", International Journal of Engineering and technology.
4. AnkitaDewan and Meghana Sharma, "Prediction of Heart Disease using a Hybrid Technique in Data Mining Classification", 2015, 2nd International Conference on Computing for Sustainable Global Development, IEEE.
5. Theresa Princy, R and J. Thomas, "Human Heart Disease Prediction system using Data Mining Techniques", 2016 International Conference on Circuit, Power and Computing Technologies[ICCPCT], IEEE.
6. Ahmad Mohawish, RaginiRathi, and VibhanshuAbhishek, "Predicting Coronary Heart Disease Risk Using Health Risk Assessment Data", SSH 2015: The Third International Workshop on Service Science for E-Health, IEEE.
7. Raid Lafta, YanLi and Vincent S. Tseng, "An Intelligent Recommender System based on Short Term Risk Prediction for Heart Disease patients", 2015, IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.
8. M.A. Jabbar, B.L. Deekshatulu and Priti Chandra, "Computational Intelligence Technique for early diagnosis of Heart Disease", 2015, IEEE International Conference on Engineering and Technology (ICETECH), 20 March 2015, Coimbatore, TN, India.
9. Syed Umar Amin, KavithaAggarwal and Dr.Rizwan Beg "Genetic Neural Network based data Mining in prediction of Heart disease Using risk factors", proceedings of 2013 IEEE conference on information and Communication Technologies(ICT 2013).
10. B.V. Bajju and R.J Remy Janet, "A survey on Heart disease diagnosis and prediction using Naïve Bayesin Data Mining", International journal of current engineering and Technology, volume 5 no 2, April 2015.
11. J.E.Nalavade, M.L.Gavali, N.D. Gohil and S.C. Jamale, "Implementing Heart Attack prediction System using Data Mining and Artificial Neural Network", International journal of Current Engineering and Technology, Vol 4, No 3, June 2014.
12. P.Priyanga and Dr.Naveen N.C, "Mining Health Data using Weighted Approach", Communication on Applied Electronics(CAE)-ISSN: 2394-4714, Foundation of Computer Science FCS, New York, USA, Vol 5- No 10, Sep 2016-www.caeaccess.org.
13. P.Priyanga and Dr.NaveenN.C, "Literature Review: Web mining techniques in Health Care applications", International Journal of Computer Engineering Applications, ICCSTAR-2016, Special issue May 16.
14. Sana Bharti and Dr.Shaliendra Narayan Singh, "Analytical Dstudy of Heart disease prediction comparing with different algorithms", International Conference on Computing, Communication and Automation(ICCCA2015), IEEE.