ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

Exploration Risk Factors and Prediction of Heart Disease Using Mining Techniques

Sowmiya.T[1], Dr. V. Sai Shanmuga Raja[2],

M.E. Computer Science and Engineering, Department of Computer Science and Engineering, Shanmuganathan Engineering College, Thirumayam Taluk, Arasampatti, Pudukkottai, TamilNadu, India.

Assistant Professor, Department of Computer Science and Engineering, Shanmuganathan Engineering College, Thirumayam Taluk, Arasampatti, Pudukkottai, TamilNadu, India.

ABSTRACT: The main aim is to predict the Heart Diseases based on the parameters, the analysis of high risk factors of developing heart diseases are identified using Intelligent Support Vector [ISV] Algorithm with rule mining techniques. The World Health Organization estimates that by 2030 there will be approximately 350 million young people (below 30 to 40 years) with heart diseases associated with renal complications, stroke and peripheral vascular disease. Heart disease is most common in present era. The treatment cost of heart disease is not affordable by most of the patients. So we can reduce this problem by a Heart Disease Prediction System (HDPS). It is helpful for earlier diagnosis of heart disease. Data mining techniques are used for the construction of HDPS. In health care field some systems use large healthcare data in varied forms such as images, texts, charts and numbers. Our aim is to analyze the risk factors and system conditions to detect heart disease early. Using effective methods to identify and extract key information that describes aspects of developing a prediction model, sample size and number of events, risk predictor selection. Using the new algorithm called Intelligent Support Vector [ISV], we can easily identify the heart disease with various attributes and risk factor specifications. Based on these parameters, the analysis of high risk factors of developing heart disease is identified using mining principles. Use of data mining algorithms will result in quick prediction of disease with high accuracy.

KEYWORDS: Prediction, Heart Disease, HDPS, Random Forest, Naïve Bayes, Health Records

I. INTRODUCTION

We propose a clinical application of mining techniques to identify sets of abnormal conditions (and the patient subpopulations who suffer from these conditions) that imply significantly increased risk of heart diseases. Intelligent Rule mining principles and Intelligent Support Vector [ISV] Algorithm on this approach to produce extensive results based on risk factor analysis and symptoms specified. The computer aided systems help the doctor as a tool for predicting and diagnosing heart disease. The medical field is dealing with huge amount of data regularly as well as handling that large data by traditional way may affect the results. Our main contribution is a comparative evaluation of these extended summarization techniques that provides guidance to practitioners in selecting an appropriate algorithm for a similar problem.

Heart is one of the important organs in blood circulatory system of all living organism. There are many elements which make problems to heart. They are smoking, poor eating methodology, high pulse, cholesterol and high blood pressure etc. The diagnosis of heart disease in earlier stage is a challenging problem for the medical industry. Data mining based heart disease prediction system can help in determining the heart disease during early stages. The prediction system helps to reduce the high risk of heart disease. Prediction is done based on the current data given to the system. For building Heart Disease Prediction System we use Microsoft Visual Studio tool.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

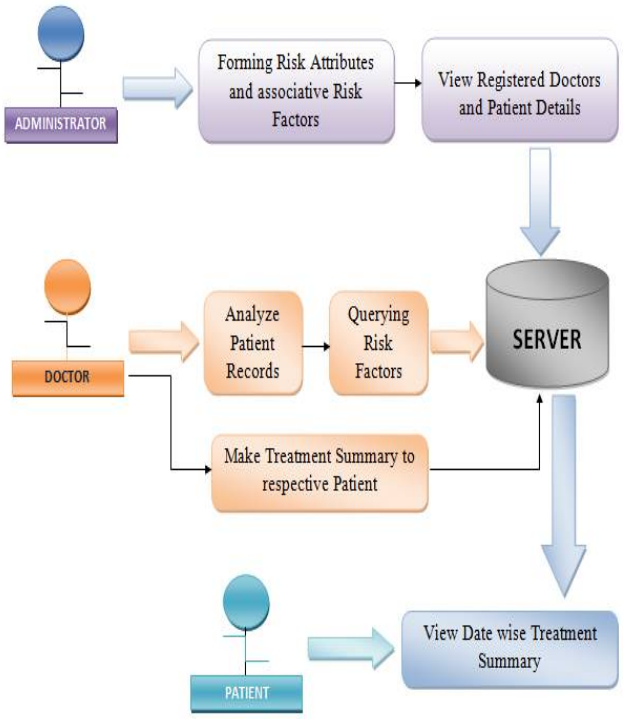

Fig.1 Proposed System Architecture

Diagnosis of heart disease is a complicated task in medical field. So it is needed to develop an efficient disease prediction system for the earlier detection of disease. One of the earliest systems for heart disease detection was proposed by Meghna Sharma et al. and they propose a hybrid technique in data mining for heart disease prediction. Here a prototype which can extract unknown data related with heart disease from a past heart disease database record is developed. They put an idea of hybrid technique methodology which can be implemented in future to have accuracy of almost 99% or with least error. In past researches, authors analyses various papers on heart on using different data mining technologies.

Also they make comparative study on the performance of three classifiers like Naïve Bayesian classifier, Decision trees and Probabilistic Neural Network (PNN). The analysis showed that artificial neural networks gave accuracy of 94.6 percentages in heart disease prediction. In earlier systems, author developed an intelligent Heart Disease Prediction System. This system is builds using Naive Bayes algorithm and it also uses a smoothing technique (Jelinek mercer smoothing) to improves the performance. Here the system is proposed using Cleveland heart disease database as the input dataset. Each attribute of the dataset were fed to the Naive Bayesian classifier and it produces the prediction results based on the classification process.

We can conclude that efficiency can be improved with the use of smoothing technique. This model could answer complex queries which traditional decision support systems cannot. In earlier systems, Bhuvaneswari Amma proposes a system using genetic algorithm and neural network, which is helpful for cardiovascular disease prediction. Observed demerits of this system are: less accuracy and no extraction of hidden data.

II. EXISTING APPROACHES – A SUMMARY

Heart Disease index is in essence a predictive model that assigns a score to a patient based on their estimated risks of suffering. Collins et al. conducted an extensive survey of Heart Disease, which indices describing the risk

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

factors and the modeling technique that these indices utilized. Lots of past classification approaches such as SVM, Neural networks, Bayesian Classification and so on are derived but all are facing struggles in certain limits. However all these methods which is developed in past systems fails to maintain the risk attributes and visualizing the summary to patients. Association Rules are used to indexing the disease summary, but the resulting does not constitute a Heart Disease index because the study does not designate a particular outcome of interest. And they do not assess or predict the risk of Heart Disease in patients, but they discovered some significant associations between diagnosis codes. The past system contains several disadvantages, some of them are listed below: (a) Comparatively Low in Performance because of classical MapReduce and Classification techniques, (b) Because of traditional classification approaches system requires large storage space to maintain the risk factor and patient records, (c) Difficult to Search the medicinal records in a single server, (d) Maintenance cost is high and (e) Low level of efficiency.

III. PROPOSED SYSTEM SUMMARY

In our proposed system, we extend our previous study by incorporating a wide variety of predictors alongside the diagnosis codes. We use the advanced methodology of combining survival analysis and distributional rule mining techniques. In this approach, we review the basic concepts underlying this methodology: survival analysis, rule mining and Intelligent Support Vector [ISV]. Disease Investigation takes place when we obtain full information about a patient and system dynamically classifies and generates the summary regarding patient health. For example, if a patient drops out of the weight level, we may not know that he/she is suffered, but with our approach we can easily identify and get back soon from the disease with the help of earlier predictions based on risk factor analysis and symptom attributes.

The ability to use such partial information and the ability to take time into account are the key characteristics of survival analysis making it a mainstay technique in clinical research. The proposed system contains several advantages, some of them are listed below: (a) Timing consumption is low for order processing, (b) Maintenance cost is comparatively low, (c) High accuracy to retrieve medicinal details from the hospital server, (d) Quicker order processing and (e) Provide the efficient authentication and authorization process to achieve privacy and security to the system.

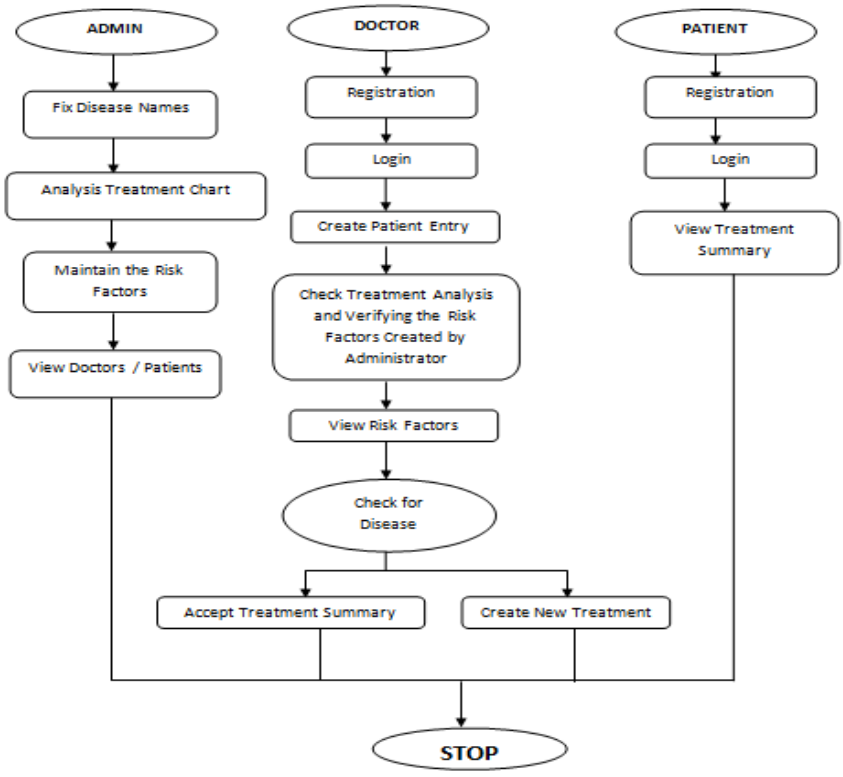

Fig.2 System Flow Diagram

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

IV. SYSTEM IMPLEMENTATION

The proposed system is designed with lots of purposeful modules that are described in detail:

A. Identify Patients Risk Attributes

In response to the pressing need to identify patients at high risk of heart disease early, numerous diabetes risk indices (risk scores) have been developed. Some of these indices gained acceptance in clinical practice and are used as guidance in treatment: patients presenting high risk scores are treated more aggressively. These scores only provide a quantification of the risk; they are not suggestive of the factors that may have caused the elevation of the risk. Moreover, these scores utilize individual risk factors in an additive fashion without taking interactions among them into account. Heart Disease is the major physical syndrome, which is a constellation of diseases including hyperlipidemia (elevated triglyceride and low HDL levels), hypertension (high blood pressure) and and central obesity (with body mass index exceeding 30 kg/m2) and so on. These diseases interact with each other, with cardiac and vascular diseases and thus understanding and modeling these interactions is important.

B. Applying Rule Mining Techniques

Rule Mining Techniques such as Association rules, Apriori and so on are implications that associate a set of potentially interacting conditions with elevated risk. The use of association rules is particularly beneficial, because in addition to quantifying the Heart Disease risk, they also readily provide the physician with a “justification”, namely the associated set of conditions. This set of conditions can be used to guide treatment towards a more personalized and targeted preventive care or Heart Disease management. While association rules themselves can be easily interpreted, the resulting rule sets can sometimes be very large, eroding the interpretability of the rule set as a whole. Especially, in this work, we consider a rich set of risk factors, namely co-morbid diseases, laboratory results, medications and demographic information that are commonly available in Electronic Medical Record (EMR) systems.

C. Symptoms And Attribute Summarization

A number of successful rule mining set summarization techniques have been proposed but no clear guidance exists regarding the applicability, strengths and weaknesses of these techniques. The focus of this manuscript is to review and characterize four existing association rule summarization techniques and provide guidance to practitioners in choosing the most suitable one. A common shortcoming of these techniques is their inability to take Heart Disease risk–a continuous outcome into account. In order to make these techniques more appropriate, we had to minimally modify them: we extend them to incorporate information about continuous outcome variables.

D. Disease Follow-Up Strategies

In this module we are concerned with two types of events: progression to Heart Disease and last follow-up. We define a status variable indicating whether a patient progressed to Heart Disease at any time during the study and we also define follow-up, which is the time from the beginning of the study to progression to Heart Diseases or last follow-up, whichever occurred earlier. Last follow-up denotes the last encounter when data is available about the patient. Note that patients who progressed to Heart Disease are no longer followed as far as survival models are concerned.

V. LITERATURE SURVEY

In the year of 2010, the authors "Ankita Dewan and Meghna Sharma" proposed a paper titled “Prediction of Heart Disease Using a Hybrid Technique in Data Mining ClassifIcation", in that they described such as: Heart disease prediction is treated as most complicated task in the field of medical sciences. Thus there arises a need to develop a decision support system for detecting heart disease of a patient. In this paper, we propose efficient genetic algorithm hybrid with the back propagation technique approach for heart disease prediction.

Today medical field have come a long way to treat patients with various kind of diseases. Among the most threatening one is the Heart disease which cannot be observed with a naked eye and comes instantly when its limitations

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

are reached. Bad clinical decisions would cause death of a patient which cannot be afforded by any hospital. To achieve a correct and cost effective treatment computer-based and support Systems can be developed to make good decision. Many hospitals use hospital information systems to manage their healthcare or patient data. These systems produce huge amounts of data in the form of images, text, charts and numbers. Sadly, this data is rarely used to support the medical decision making. There is a bulk of hidden information in this data that is not yet explored which give rise to an important query of how to make useful information out of the data. So there is necessity of creating an excellent project which will help practitioners predict the heart disease before it occurs.

The main objective of this paper is to develop 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.

In the year of 2015, the authors "Monika Gandhi and Dr. Shailendra Narayan Singh" proposed a paper titled “Predictions in Heart Disease Using Techniques of Data Mining", in that they described such as: As huge amount of information is produced in medical associations (healing facilities, therapeutic focuses) yet this information is not properly utilized. The health care system is "data rich" however "knowledge poor ". There is an absence of successful analysis methods to find connections and patterns in health care data. Data mining methods can help as remedy in this circumstance. For this reason, different data mining techniques can be utilized. The paper intends to give details about various techniques of knowledge abstraction by using data mining methods that are being used in today's research for prediction of heart disease. In this paper, data mining methods namely, Naive Bayes, Neural network, Decision tree algorithm are analyzed on medical data sets using algorithms.

In the year of 2014, the authors "Ms.Rupali and R.Patil" proposed a paper titled “Heart Disease Prediction System using Naïve Bayes and Jelinek-mercer smoothing", in that they described such as: Heart disease is most common in present era. The treatment cost of heart disease is not affordable by most of the patients. So we can reduce this problem by a Heart Disease Prediction System (HDPS).It is helpful for earlier diagnosis of heart disease. Data mining techniques are used for the construction of HDPS. In health care field some systems use large healthcare data in varied forms such as images, texts, charts and numbers .But this datas are hardly visited and are not mined. This problem can be avoided by introducing HDPS. This system would enhance medical care and it can also reduce the costs.

The system can handle complex queries for detection of heart disease and thus help to make intelligent medical decisions. This paper proposes a HDPS based on three different data mining techniques. The various data mining methods used are Naive Bayes, Decision tree (J48), Random Forest and WEKA API. The system can predict the likelihood of patients getting a heart disease by using medical profiles such as age, sex, blood pressure, cholesterol and blood sugar. Also, the performance will be compared by calculation of confusion matrix. This can help to calculate accuracy, precision, and recall. The overall system provides high performance and better accuracy.

VI. EXPERIMENTAL RESULTS

The following figure shows the home page of the administrator.

Fig.3 Administrator Home Page

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

The following figure illustrates the Risk Factor Updation Details of the proposed system.

Fig.4 Risk Factor Updation

The following figure illustrates the Risk Attribute Master Details of the proposed system.

Fig.5 Risk Attribute Master

VI. CONCLUSION

Heart disease is the leading cause of death for both men and women. Know the warning signs and symptoms of a heart attack so that you can act fast if you or someone you know might be having a heart attack. The chances of survival are greater when emergency treatment begins quickly. This system mainly focuses on the experimental analysis of various approaches of heart attack disease prediction researches and proves the efficiency via proposed results of ISV.

REFERENCES

[1] Ankita Dewan and Meghna Sharma "Prediction of Heart Disease Using a Hybrid Technique in Data Mining ClassifIcation" IEEE2010
[2] Majali J, Niranjan R, Phatak V,Tadakhe O.Data mining techniques for diagnosis and prognosis of cancer. International Journal of Advanced Researchin Computer and Communication Engineering.2015;4(3):613–6.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

[3] Ms.Rupali R.Patil "Heart Disease Prediction System using Naïve Bayes and Jelinek-mercer smoothing" IJARCCE 2014
[4] Monika Gandhi and Dr. Shailendra Narayan Singh "Predictions in Heart Disease Using Techniques of Data Mining" International Conference Conference on Futuristic trend in ComputationalAnalysis and Knowledge Management (ABLAZE-2015)
[5] Bhuvaneswari Amma N.G.," Cardiovascular Disease Prediction System using Genetic Algorithm and Neural Network" 2014
[6] Chaitrali S. Dangare et. al.,"Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", (IJCA) (0975 – 8887), Vol. 47, No. 10, June 2012, page no. 44-48.
[7] S. Vijiyarani et. al., "An Efficient Classification Tree Technique for Heart Disease Prediction", (ICRTCT - 2013) Proceedings published in (IJCA) (0975 – 8887), 2013, page no. 6-9.
[8] Y.E.Shao, C.-D. Hou, and C.-C. Chiu,"Hybrid intelligent modelling schemes for heart disease classification, "Applied Soft Computing, 2014.
[9] Y. Xing, J. Wang, Z. Zhao, and Y. Gao, "Combination data mining methods with new medical data to predicting outcome of coronary heart disease," pp. 868–872, 2007.
[10] Guru, N., Anil, D., Navin, R., Decision Support System FoR Heart Disease Diagnosis Using Neural Network.Delhi Business Review.
[11] Shouman M, Turner T, Stocker R. Using decision tree for diagnosing heart disease patients. Proceedings of the 9th Australasian Data Mining Conference(AusDM'11); Ballarat, Australia. 2011. p. 23–30.
[12] RaniKU.Analysis of heart diseases dataset using neural network approach.IJDKP. 2011.
[13] Y. Ji, P. Xu, and Y. Ye, "HVAC terminal hourly end-use disaggregation in commercial buildings with Fourier series model," Energy and Buildings, 2015.
[14] B. M. Brentan, E. Luvizotto Jr, M. Herrera, J. Izquierdo, and R. Prez-Garca, "Hybrid regression model for near real-time urban water demand forecasting," Journal of Computational and Applied Mathematics, 2016.
[15] F. K. Odan, and L. F. R. Reis, "Hybrid water demand forecasting model associating artificial neural network with Fourier series," Journal of Water Resources Planning and Management, vol. 138, no. 3, pp. 245-256, 2012.
[16] K. Samiee, P. Kovcs, and M. Gabbouj, "Epileptic seizure classification of eeg time-series using rational discrete short-time Fourier transform," IEEE Transactions on Biomedical Engineering, vol. 62, no. 2, pp. 541-552, 2015.
[17] P. Kovacs, K. Samiee, and M. Gabbouj, "On application of rational discrete short time Fourier transform in epileptic seizure classification," IEEE Transactions on Biomedical Engineering, pp. 5839-5843, 2014.
[18] P. Mohanty, S. Ahmad, H. Antia, K. Arunbabu, A. Chandra, S. Dugad, S. Gupta, B. Hariharan, Y. Hayashi, and P. Jagadeesan, "Fast Fourier transform to measure pressure coefficient of muons in the GRAPES-3 experiment," Astroparticle Physics, vol. 79, pp. 23-30, 2016.