

Comparative Study of Predicting Heart Disease By Means Of Data Mining

C.Sowmiya¹, Dr.P. Sumitra²

¹Ph.D Research Scholar, Department of Computer Science,

²Professor, Department of Computer Science,

^{1,2}Vivekanandha College of Arts and Sciences for Women (Autonomous),Elayampalayam.

ABSTRACT

At the present time all the activities are done in the world through internet. The use of computer in the fields of medicine are highly improved. The computerized hospital includes the activities like treatment of illnesses, maintain the patient's information up-to-date that leads to handle huge amount of data regularly. It is very difficult to handle the large amount of data for predicting the heart disease. Data mining approach is a good way for predicting the heart disease at minimum effort. This paper analyses the existing prediction system and discussing the various disputes on the existing system. The result of this paper helps the health consultants to diagnose the disease in less time and predict the probable problems well in advance and save the patient life.

Keywords: Data mining, Heart diseases, classification

I. INTRODUCTION

Data mining turn into the ultimate method for finding the practical solutions to the day to day problems, and the health care is no exception to this. Most of the data mining methods are developed to help clinicians for making better decisions about the patient treatment purposes. nowadays a data mining techniques are mostly used in the areas such as healthcare organizations, health informatics, epidemiology, patient care and monitoring systems, assistive technology, large-scale image analysis to information extraction and automatic identification of unknown classes. The goal of this paper is to analyze several datamining techniques that are existed in recent years for the diagnosis of heart disease. Many researchers used datamining techniques in the diagnosis of diseases such as tuberculosis, diabetes, cancer and heart disease, in which several data mining techniques are used in the diagnosis of heart disease such as KNN, Neural Networks, and Bayesian classification. Classification based on clustering, Decision than just heart attacks. It applies to a number of conditions that affect the heart, including coronary artery disease, arrhythmias, atrial fibrillation, heart valve disease, congenital heart disease, cardiomegaly (enlarged heart), cardiomyopathy (heart muscle disease), and more. Heart disease is the number one cause of death in the United States for both men and women. Sometimes, symptoms can be subtle and go unnoticed until a major event like a heart attack occurs. Noticeable symptoms include: chest pain (angina), extreme fatigue, and shortness of breath. Certain lifestyle habits and risk factors contribute to heart disease. Some risk factors—like age and gender—can't be controlled. However, others can be. To keep your heart healthy, it's a good idea to lower your blood pressure, eat a high-fiber, low-fat diet, exercise regularly, manage your stress, and quit smoking.

A. Types of heart diseases:

Heart disease is a word used to describe many different conditions affecting the heart. Coronary heart disease is a common type of heart disease. This condition results from a buildup of plaque on the inside of the arteries, which reduces blood flow to the heart and increases the risk of a heart attack

Tree, Genetic Algorithm, Naive Bayes, Decision tree, WAC which are showing accuracy at different levels. Using medical profile such as age, sex, blood pressure and blood sugar we can easily predict the likelihood of patients getting heart disease.

II. HEART DISEASE

Heart disease labels that is a series of conditions that affects the heart. Diseases under the heart disease umbrella include blood vessel diseases, such as coronary artery disease; heart rhythm problems. The term "heart disease" is often used interchangeably with the term "cardiovascular disease." Cardiovascular disease pointed or choked blood vessels that can leads several problem like a heart attack, chest pain (angina) or stroke. In most of the countries death is caused by the heart diseases that is surveyed by WHO (World Health Organization) and the CDC. The adults of the countries UK, USA, Canada and Australia be distressed from heart disease. The term "heart disease" refers to more

and other heart complications. Other forms of heart disease include:

- irregular heartbeat (arrhythmias)
- congenital heart defects
- weak heart muscles (cardiomyopathy)
- heart valve problems
- Heart infections
- cardiovascular disease

B. Heart disease statistics

Approximately 610,000 people die from heart disease in the United States every year, according to the Centers for Disease and Control Prevention (CDC). It's the leading cause of death in both men and women. Coronary heart disease is the deadliest of all heart diseases, just as it's the most common form. The Heart Foundation estimates 380,000 related deaths per year. The symptoms of heart disease vary between gender. Some are more obvious in men, who made up more than half of all heart disease-related deaths in the United States in 2009, according to the CDC. According to The Heart Foundation, 1 in 3 women die of heart disease every year in the United

States. In 90 percent of these cases, women had at least one preventable risk factor.

C. Symptoms of heart disease

Heart disease is often called a “silent killer.” Your doctor may not diagnose the disease until you show signs of a heart attack or heart failure. Symptoms of heart disease vary depending on the specific condition. For example, if you have a heart arrhythmia, symptoms may include:

- a fast or slow heartbeat
- dizziness
- lightheadedness
- chest pains
- shortness of breath

Symptoms of a congenital (present at birth) heart defect may include skin discoloration, such as a bluish or pale color. You may also notice swelling in your legs and stomach. You might become easily tired or have shortness of breath shortly after beginning any type of physical activity.

D. Heart Failure

Heart failure happens when the heart isn't pumping enough blood to meet your body's needs. While many people believe the misconception that heart failure means an individual is about to die or that their heart has stopped, this is not true. Heart failure simply indicates that the heart is not squeezing as well as it should. It usually does not occur suddenly but gradually worsens over the time. Heart failure can be caused by:

- Cardiomyopathies
- Coronary Artery Disease
- Diabetes
- Diseases of the Heart Valves
- Heart Defects present at Birth
- High Blood Pressure
- Lung Disease such as Emphysema
- Past Heart Attacks

III. DATA MINING

There is an unstoppable growth in the amount of electronic health records or EHRs being collected by healthcare facilities. It has been the norm for nurses to take responsibility in handling patient data input that was traditionally recorded in paper-based forms. Accuracy is extremely important when it comes to patient care and computerizing this massive amount of data enhances the quality of the whole system. Data mining has been used to uncover patterns from the large amount of stored information and then used to build predictive models. Improving the quality of patient care and reducing healthcare costs are the ideal goals of many programs. Data mining has helped these programs succeed.

Benefits of Data Mining in the Healthcare Industry

While other solutions might favor healthcare providers or insurance companies, data mining benefits everyone concerned, from healthcare organizations to insurers to patients. Patients receive more affordable and better healthcare services. This happens when healthcare officials use data mining programs to identify and observe high-risk patients and J. Shreve, H. Schneider, O. Soysal proposed a paper “A methodology for comparing classification methods through the assessment of model stability and validity in variable selection. Decision Support Systems “. In this paper the author compare the performance of classification methods by using Monte Carlo simulations and illustrates that the variable selection process is integral in comparing methodologies to ensure

chronic diseases and design the right interventions needed. These programs also reduce the number of claims and hospital admissions, further streamlining the process. HealthCare providers use data mining and data analysis to find best practices and the most effective treatments. These tools compare symptoms, causes, treatments and negative effects and then proceed to analyze which action will prove most effective for a group of patients. This is also a way for providers to develop the best standards of care and clinical best practices.

Insurers are now able to better detect medical insurance abuse and fraud because of data mining. Unusual claims patterns are easier to spot with this tool and it can identify inappropriate referrals and fraudulent medical and insurance claims. When insurers reduce their losses due to fraud, the cost of health care also decreases.

Healthcare facilities and groups use data mining tools to reach better patient-related decisions. Patient satisfaction is improved because data mining provides information that will help staff with patient interactions by recognizing usage patterns, current and future needs, and patient preferences.

IV LITERATURE SURVEY

Rovina D Britto, Anuradha Srinivasaraghavan in their paper, “Comparative Analysis of Accuracy on Heart Disease Prediction using Classification Methods “ concern with four algorithms like Naïve Bayes, Decision tree , K nearest neighbor and Support vector machine. In this paper the author suggest 3 phases. In that, the size of the data set will be increased by using first two algorithms and predict the heart disease by using logistic regression [1].

Lemke F, Mueller J-A. proposed “Medical data analysis using self-organizing data mining technologies”. In this paper the author construct the decision tree by analyzing every medical procedure or medical problem. The problem of this paper to collect large amount of data for constructing decision tree.[2]

K.Srinivasan proposed “Application of Data Mining Technique in Healthcare and Prediction of Heart Attacks”. The author suggest four algorithms such as Rule based, Decision tree, Naïve Bayes and Artificial Neural Network to the massive Volume of healthcare data. The author use Tanagra data mining tool for their data analysis. The author analysis 3000 instances with fourteen different attributes. [3]

Latha Parthiban presented a paper, “ model on basis of Coactive Neuro-Fuzzy Inference System (CANFIS) for prediction of heart disease”. In this model the author compressing the different techniques that includes the neural network adaptive capabilities, the fuzzy logic qualitative approach and further integrating with genetic algorithm. [4]

Lei Yu and Huan Liru presented a paper “Accurate and efficient classification based on multiple association rules”. The author proposed a novel method for their classification and named as predominant correlation. The proposed a fast filter method can identify relevant features without pair wise correlation analysis. [5]

minimal bias, enhanced stability, and optimize performance. [6]

P. Chandra, M. Jabbar, and B. Deekshatulu, offered “Prediction of Risk Score for Heart Disease using Associative Classification and Hybrid Feature Subset Selection”. Association rule algorithm is used to determine relations between attributes of data instances.

The classification algorithm is used to predict the classes in the patient dataset. Feature selection algorithm determines attributes which helps in predicting heart diseases. [7]

Usha Rani proposed “Disease Dataset using neural network approach”. In this paper the author suggest artificial neural network that combines forward and backward propagation algorithm. The experiment is conducted by single and multilayered neural network models. Parallelism is implemented to speed up the learning process at each neuron in all hidden and output layer. [8]

AnkitaDewan proposed ,” Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification” .In this paper the author suggest combination of data mining algorithm for predicting the heart disease. The result of this paper concludes that the neural network is best among all the classification techniques for a non-linear data. The author suggest that the Back Propagation algorithm is the best classifier of Artificial Neural Network which is a common method of training. But it has main drawback of held in a local minima. [9]

V. CONCLUSION

This paper presented a review of disease prediction for healthcare system using data mining techniques. Medical related information’s are volumetric in nature and it could be derived from different birthplaces which are not entirely applicable in a feature. In this work, we have performed a literature survey on various papers. In future, we are planning to propose an effective disease prediction system to predict the heart disease with better accuracy using different data mining classification techniques such as Naïve Bayes and Decision Tree.

VI. REFERENCES

[1] RovinaDbritto, AnuradhaSrinivasaraghavan, “Comparative Analysis of Accuracy on Heart Disease Prediction using Classification Methods”

[2].Lemke F, Mueller J-A. “Medical data analysis using self-organizing data mining technologies”,SystemsAnalysisModeling Simulation. 2003; 43(10):1399–408.

[3]. Parthiban L, Subramanian R. Intelligent heart disease prediction system using CANFIS and genetic algorithm. International Journal of Biological, Biomedical and Medical Sciences. 2008; 3(3).

[4]. Li W, Han J, Pei J. CMAR: accurate and efficient classification based on multiple association rules. Proceedings of 2001 International Conference on Data Mining; 2001.

[5]. Shreve J, Schneider H, Soysal O. A methodology for comparing classification methods through the assessment of model stability and validity in variable selection. Decision Support Systems. 2011; 52:247–57.

[6]0. Wanga T, Huang H, Tian S, Xu J. Feature selection for SVM via optimization of kernel polarization with Gaussian, ARD kernels.. Expert Systems with Applications. 2010; 37:6663–8.

[7]. Patil SB, Kumaraswamy YS. Extraction of significant patterns from heart disease warehouses for heart attack prediction. International Journal of Computer Science and Network Security (IJCSNS). 2009; 9(2):228–35.

[8]. Deepika N.,Chandrashekar K. Association rule for classification of Heart Attack Patients. International Journal of Advanced Engineering Science and Technologies. 2011; 11(2):253–57.

[9]. Srinivas K, Rani KB, Govrdhan A. Application of data mining techniques in healthcare and prediction of heart attacks. International Journal on Computer Science and Engineering. 2011; 2(2):250–5.

[10]P. Chandra, M. Jabbar, and B. Deekshatulu, “Prediction of Risk Score for Heart Disease using Associative Classification and Hybrid Feature Subset Selection,” in 12th International Conference on Intelligent Systems Design and Applications (ISDA), 2012, pp. 628–634.

[11]Usha. K Dr, “Analysis of Heart Disease Dataset using neural network approach”, IJDKP, Vol 1(5), Sep 2011.

[12]AnkitaDewan, Meghna Sharma,” Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification”, 2nd International Conference on Computing for Sustainable Global Development IEEE 2015 pp 704-706.

[13].www.kmd.ovgu.de/kdd15_medical_mining_tutorial.html

[14]<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2916206/>

[15] www.medicalnewstoday.com/articles/237191.php

[16]<http://www.healthline.com/health/heart-disease/type>