

Prediction of Heart Disease through Data mining Techniques – A survey

Anushya.R¹, Hiindhushri.S², Krishnendhu.R³, Lazar Galen Perez.W⁴, Kirubha.M⁵

^{1,2,3,4}UG Scholar, ⁵Assistant Professor
Department of Computer Science and Engineering,
Sri Ramakrishna Institute of Technology, India

Abstract: Heart malfunction is the leading disease worldwide. The detection of risk factors and their progress is complicated. The doctors and experts are not available in proportion to the population. In order to assist the physicians by developing an intelligent frame work for diagnosing the disease in heart is one of the essential goals of the researchers. The improved decision tree as C 5.0 is used for the effective extraction of the feature from data set for analysis. Using techniques like Naive Bayes, Neural Network, KNN algorithm and Decision Trees - an Intelligent Heart Disease Prediction System is created. In this paper, the prediction is done accurately using the above mentioned algorithms.

Keywords: Improved decision tree - C 5.0, Naive Bayes, Neural Network, KNN Algorithm, Decision Trees, Heart Disease Prediction System

Citation: Anushya.R, Hiindhushri.S, Krishnendhu. R,Lazar Galen Perez.W, Kirubha.M.
Article: Diabetic Retinopathy Grading System Using Machine Learning. *International Journal of Computer Science and Engineering Communications* 6(1): 1821-1824, April 2018.

Copyright © 2018 Anushya.R et al., This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION:

Data Mining is the way toward dealing with expansive data collections to recognize designs and set up connections to take care of issues through information examination. These techniques are used to explore, analyze and extract data using complex algorithms in order to discover patterns. It is used to predict future trends. Nowadays the heart disease rate is high

(i.e) life of people is at risk. Therefore detection and prediction of heart disease is done by considering various attributes and applying data mining techniques to obtain high accuracy rate.

LITERATURE SURVEY:

KaanUyar et al., describes that The World Health Organization (WHO) estimated that around 17.7 million people lost their lives from cardiovascular diseases in 2015. The Cleveland heart disease dataset was taken by The University of California Irvine (UCI). Their outcome demonstrated that 97.78% accuracy. In expansion to that root mean square blunder, the likelihood of the misclassification mistake, specificity, affectability, exactness and F-score are figured. Heart failures arise because of coronary problems, high blood pressure, diabetes, hypertension, poor diet, poor hygiene and physical inactivity[1]. The data sets are separated into two subsets for training and testing. It can be improved by adding additional attributes and may also be useful for better accuracy.

Ms.TejaswiniU.Mane. Specifies that heart is the main organ of the human body, if it can't work properly then brain will get suffered and it leads to death. Big Data is used to extract the data from the dataset. For clustering the database, improved K-means algorithm is applied for the accuracy of the Clustering Centroid. ID3 algorithm is used in the Hybrid approach for classification [2]. The decision tree is built by using the Standard Database and algorithm so that the user can pass the parameters and the status of the heart disease will be shown.

K.Sudhakar et al., suggests that by using data mining techniques Naive Bayes, Neural Network and Decision Trees, an Intelligent Heart Disease Prediction System (IHDPS) is developed. Each method has its own strength to get appropriate results. Hidden patterns and relationship is used to build this system. It is web-based, user friendly and expandable. For training and testing of data, Neural Network with back propagation algorithm is used. The hereditary calculation is utilized as a part of the revelation of abnormal state expectation rules and the found tenets are profoundly intelligible, having high predictive precision and of high intriguing quality esteems [3]. The main objective is to study the various data mining techniques available to predict the heart disease and to compare them to find the best method of prediction.

RoniShouval et al., explain that the treatment of acute ST-segment elevation myocardial infarction(STEMI) has significantly advanced in recent years. The method of data analysis that automates analytical model building is called Machine learning (ML). ML algorithms is used in data mining for knowledge discovery and predictive modeling in large datasets[4]. The feasibility and competence of ML tools for predictive modeling in complex data scenarios within the field of cardiology which can further aid clinicians in developing tools for more precise patient risk stratification.

Monika Gandhi et al., delivers that the health care systems are normally rich in data but poor in knowledge. There are no successful analysis methods to analyze patterns with the available data. Therefore various techniques of knowledge abstraction using data mining

algorithms are obtained to predict heart disease[5]. Data mining methods used are naive Bayes, neural networks, decision tree algorithm to analyze and predict heart disease. These techniques help in finding out the hidden patterns for making decisions in health care system.

Vivekanandan et al., defines that the traditional record maintenance is replaced in health care systems. The diagnosis of heart disease is carried out by selecting critical features from enormous set of available data. Modified differential algorithm is used to perform feature selection for cardiovascular disease. Prediction of heart disease is carried out using fuzzy AHP and feed forward neural networks. The accuracy is 83%. This model gives prediction of heart disease with reduced prediction time and better performance [6]. Further error elimination will be carried out to improve performance.

Purushottam et al., suggest that a framework has been developed that can proficiently find the tenets to foresee the risk level of patients in the view of given parameters about the health. Several rules are generated by the proposed systems. The execution of the framework is assessed and demonstrates that the framework has extraordinary potential in anticipating the coronary illness risk level. This model helps in effective decision making for the medical practitioner [7]. This app gives 88% accuracy. It helps both patient and advisor the early determination results.

Thomas et al., describe that heart disease has become more common these days. So by considering various factors such as age, blood pressure, pulse rate etc. Classification algorithms are applied over them for prediction of heart disease [8]. More the number of parameters considered more the accuracy rate. KNN and ID3 algorithm was used to detect the prediction of risk rate and accuracy is high by considering many attributes. The future work is increasing the accuracy using minimum number of attributes.

Meyyappan.T et al., explain that Data mining techniques are used in the field of medicine for various purposes where frequent itemsets are generated. It was first proposed for market basket analysis. Researchers proposed variations in techniques to generate frequent itemsets. Datasets of various heart disease patients are used for research. Association data mining technique is efficient for generating itemset. In US a major threat to human health is Heart Disease. Nearly 1 million people get affected annually. For every 30 sec, an individual has heart attack. For every sec, an individual dies. Algorithm for discovering maximum frequency item set reduces the number of times the database needs to be scanned [9]. K means algorithm is used to predict risky and non risky factors. The Symptoms include chest pain, shortness of breath, breathlessness, lightheadedness etc. The training datasets include 1000 patients affected by heart related disease with 19 clinical attributes.

Mamta Sharma et al., suggest that Data mining plays an effective role for uncovering new trends in healthcare organization and it is one of the richest areas of research that is more popular in health organizations [10]. The data used in mining may be in the form of free text, structured as in databases or in the form of images. It explores the utility of various decision tree and neural network algorithms to classify and predict the disease.

Carlos Ordonez describes that Association rules represent a promising technique to improve heart disease prediction. When association rules are applied on a medical data set, they produce an extremely large number of rules. Most of such rules are medically irrelevant and the time required to find them can be impractical. Each rule represents a simple predictive pattern that describes a subset of the dataset projected on a subset of attributes[11]. There is a high proportion of rules that cannot be generalized after validation on the test set. A hierarchy of perfusion measurements is required to control the rule discovery process, in order to increase or decrease sensitivity to detect sick patients with high accuracy without significantly losing sensitivity.

Raid Lafta et al., describes that The use of intelligent technologies in clinical decision making in the telehealth environment has begun to play a vital role in improving the quality of patients' lives and helping reduce the costs and workload involved in their daily healthcare. The input sequence of sliding windows based on the patient's time series data are decomposed by using the fast Fourier transformation in order to extract the frequency information[12]. A recommendation system supported by a machine learning ensemble model with the fast Fourier transformation for short-term disease risk prediction and medical test recommendation in the telehealth environment for patients suffering from chronic heart disease. The system improves the quality of clinical evidence-based decisions and help reduce the time costs incurred by the chronic heart disease patients in taking their daily medical test, whereby improving their overall life quality.

Sulabha S. Apte et al., specifies that to build Intelligent Heart Disease Prediction System that gives diagnosis of heart disease using historical heart database. In this paper, two more input attributes obesity and smoking are used to get more accurate[13]. Data mining classification techniques applied are Decision trees, Naive Bayes & Neural Networks. From this Neural Networks provides accurate results. To get more appropriate results, two more attributes i.e. obesity and smoking are used, as these attributes are considered as important attributes for heart disease.

CONCLUSION:

It is specified that different Data Mining strategies such as Ordering, Grouping and Affiliation are utilized by Medicinal Services Association to expand their capacity for settling on choice with respect to patient wellbeing. There are adequate examination assets accessible in regards. Data Mining errands which are exhibited in consequent areas with their points of interest and burdens. It provides the accuracy of prediction system using specified algorithms. The various other data sets and different types of algorithms can be considered for increasing the efficiency. The future work is to analyze the factors that are difficult for decision making with the help of attributes and to provide accurate results.

REFERENCES:

- [1] KaanUyar, Ahmed Ilhan "Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural network",ELSEVIER,2017.
- [2] Ms.TejaswiniU.Mane"Smart heart disease prediction system using Improved K-Means and ID3 on Big Data", IEEE, 2017.
- [3]K.Sudhakar, Dr. M. Manimekalai" Study of Heart Disease Prediction using Data Mining", IJARCSSE, Volume 4, Issue 1,January 2014.
- [4]RoniShouval, Amir Hadanny, NirShlomo, ZazaLakobishvili, Ron Unger, DoronZahger, Ronny Alcalai, ShaulAtar, ShmuelGobblieb, ShlomiMatetzky, Ilan Goldenberg, Roy Beigel"Machine learning for prediction of 30 day mortality after ST elevation myocardial infraction : An Acute Coronary Syndrome Israeli Survey data mining study",ELSEVIER,2017.
- [5]Monika Gandhi,Dr.Shailendra Narayan Singh "Prediction in Heart Disease Using Techniques of Data Mining", IEEE, 2016.
- [6] Vivekanandan, N.Ch SrimanNarayanaIyengar "Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease", ELSEVIER,2017.
- [7] Purushottam, Prof.(Dr.)KanakSaxena, Richa Sharma "Efficient Heart Disease Prediction System", ELSEVIER,2016.
- [8] J.Thomas, Theresa Princy.R "Human Heart Disease Prediction System using Data Mining Techniques", IEEE, 2016.
- [9] Ilayaraja M, Meyyappan T "Efficient Data Mining Method to Predict the Risk of Heart Diseases through Frequent Itemsets", ICECCS 2015.
- [10] Prof.Mamta Sharma, Farheen Khan, VishnupriyaRavichandran "Comparing Data Mining Techniques Used For Heart Disease Prediction",IRJET,2017.
- [11] Carlos Ordonez " Association Rule Discovery With the Train and Test Approach for Heart Disease Prediction" , IEEE , 2006.
- [12] Raid Lafta, JiZhang, Xiaohui Tao, Yan Li, Xiaodong Zhu, YonglongLuo and Fulong Chen "Coupling a Fast Fourier Transformation with a Machine Learning Ensemble Model to Support Recommendations for Heart Disease Patients in a Telehealth Environment" , IEEE,2016.
- [13] Sulabha S. Apte,ChaitraliS.Dangare "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications,2012.