



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 4, Issue 2)

Available online at: www.ijariit.com

Heart disease prediction using classification and feature selection techniques

Navdeep Singh

Shaheed Bhagat Singh State Technical Campus,
Ferozpur, Punjab

Sonika Jindal

Shaheed Bhagat Singh State Technical Campus,
Ferozpur, Punjab

ABSTRACT

The value of measurement age, blood pressure, weight, smoking habits, exercise, and blood serum cholesterol in predicting death and coronary cardiopathy was studied over an amount of ten years. Sixteen teams comprising 12 763 men aged 40-59 years (at the outset) from 7 countries (Yugoslavia, Finland, Italy, Kingdom of The Netherlands, Greece, USA, and Japan) were studied. the very best risk factors were found to be age, pulse blood pressure, and blood serum cholesterol concentration (related to saturated fatty acids within the diet). Variations in incidence rates couldn't be shown to be associated with characteristics of the cohorts in relative weight, smoking habits or physical activity. To design a perceptive model for heart illnesses acknowledgment using data mining strategies that are fit for enhancing the constancy of heart infections conclusion. Thereafter, we divide this data into Training and Testing Data Sets and employ Naïve Bayes technique to obtain relatively higher prediction accuracy. The primary goal of this research would be given a highly accurate prediction of Heart Disease. As we have done a combination of Genetic and Naïve Bayes Technique, the Investigation would be developed a Hybrid model of both these techniques and called it Hybrid Genetic Naïve Bayes Model for predicting high accuracy in results.

Keywords: Classification, Naïve Bayes, Heart Disease, and Predictive Analysis.

1. INTRODUCTION

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 a 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. A major challenge confronting healthcare associations i.e. hospitals, medicinal focuses are the procurement of quality services at reasonable expenses. Quality services suggest diagnosing patients accurately and overseeing medicines that are more effective. Poor clinical decisions can prompt to poor outcomes, which are therefore unsatisfactory. Healthcare organizations can reduce costs by accomplishment of computer-based data and/or decision support systems. Healthcare services data is very huge as it incorporates patient records, resource management information and updated information. Human services associations must have the capacity to break down information. Treatment records of many patients can be stored away in a computerized way; furthermore, data mining methods may help in finding out a few vital and basic inquiries related to healthcare organizations. Clinical choices are frequently made focused on doctors' instinct and experience instead of on the knowledge-rich information covered up in the database. This practice prompts undesirable biases, blunders and unnecessary medicinal expenses, which influence the quality of services given to the patients. Wu, et al proposed that combination of clinical choice backing with computer-based patient records could decrease medical errors, enhance the safety of patients, lessening undesirable practice variety, and enhance patient outcome. This suggestion is guaranteeing as the data demonstrating and analysis tools for example data mining, have the possibility to create a knowledge-rich environment which can help to essentially enhance the nature of clinical decisions.

Electrocardiogram

ECG (Electrocardiogram) is an interpretation tool that tells the electrical activity of heart recorded electrode by skin [1]. The morphology and pulse rate emulate the internal organ health of human heartbeat [18]. It's a not invasive technique that's used for

the signal is measured on the surface of a person's body, that is employed in the identification of the heart diseases [19]. The amplitude and period of the P-QRS-T wave contain helpful data regarding the nature of illness afflicting the heart. The electrical wave is because of depolarization and repolarization of Na⁺ and k⁻ ions within the blood [9]. The electrocardiogram signal provides the subsequent data about a human heart [10]:

- Heart rhythm and conductivity disturbances extent and site of cardiac muscle ischemia changes in solution concentrations drug effects on the heart.

Data Mining

Due to a good convenience of giant quantity info and a requirement to convert this offered huge quantity of knowledge to helpful information necessitates the utilization of knowledge mining techniques. Data processing and KDD (knowledge discovery in the database) became common in recent years [13]. the recognition of data mining and KDD (knowledge discovery in database) shouldn't be a surprise since the dimensions of the info collections that are offered are so much giant to be examined manually and even the ways for automatic data analysis supported classical statistics and machine learning usually face issues once process large, dynamic information collections consisting of complicated objects [10].

Data Mining is cored a part of knowledge Discovery database (KDD). Many of us treat data processing as an equivalent word for KDD since it's a key a part of KDD process.

HEART disease

A significant challenge facing tending organizations (hospitals, medical centers) is that the provision of quality services at reasonable prices[1]. Quality service implies diagnosing patients properly and administering treatments that are effective. Poor clinical choices will result in unfortunate consequences that are so unacceptable. Hospitals should conjointly minimize the price of clinical tests. they'll accomplish these results by using acceptable computer-based data and/or call support systems[21][20].

The heart is that the necessary a part of our body. Life is itself obsessed on economical operating of heart. If the operation of the heart isn't correct, it'll have an effect on the opposite body components of human-like the brain, kidney etc. heart illness may be a disease that affects on the operation of the heart. There is a range of things that will increase the risk of cardiovascular disease [12].

Some of them are listed below:

- The family history of heart disease
- Smoking
- Cholesterol
- High blood pressure
- Obesity
- Lack of physical exercise

2. LITERATURE SURVEY

Gandhi, Monika et al. Authors have studied 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.

Moloud Abdar, Sharareh R. Niakan Kalhori et al. Authors have compared different techniques. Authors have purposed the comparison of different data mining algorithm on the prediction of heart diseases. This work applied and compared data mining techniques to predict the risk of heart diseases. After feature analysis, models by five algorithms including C5.0, Neural Network, Support Vector Machine (SVM), K-Nearest Neighborhood (KNN) and Logistic Regression, developed and validated. C5.0 Decision tree has been able to build a model with the greatest accuracy 93.02%, KNN, SVM, Neural Network have been 88.37%, 86.05%, and 80.23% respectively. Produced results of the decision tree can be simply interpretable and applicable; different clinical practitioner can understand their rules easily.

G. Purusothaman et al. Authors have studied Comparison of classification techniques in Data mining to find the best technique for creating risk prediction model of heart disease at minimum effort. This paper provides a quick and easy understanding of various prediction models in data mining and helps to find the best model for further work. This is unique approach because various techniques listed and expressed in a bar chart to understand accuracy level of each. These techniques are chosen based on their efficiency in the literature. In previous studies of different researcher expressed their effort on finding the best approach for risk prediction model and here we found the best model by comparing those researcher's findings as a survey. This survey helps to understand the recent techniques involved in risk prediction of heart disease at classification in data mining. Survey of relevant data mining techniques, which are involved in risk prediction of heart disease, provides best prediction model as hybrid approach comparing with the single model approach.

SHABANA ASMI P et al. Authors have worked to make the system more accurately predict the presence of heart disease with added attributes of the disease and using association rules. Final results show that association rule implemented on dataset produces better accuracy.

Sunayan Bandyopadhyay, Julian Wolfson et al. Author's approach can lead to better predictive performance than the Cox proportional hazards model (i.e., a regression-based approach commonly used for censored, time-to-event data) or a Bayesian

network with ad hoc approaches to right-censoring. Our techniques are motivated by and illustrated on data from a large U.S. Midwestern health care system.

Javier Andreu-Perez, Carmen C. Y. Poon et al. Authors provide an overview of recent developments in big data in the context of biomedical and health informatics. It outlines the key characteristics of big data and how medical and health informatics, translational bioinformatics, sensor informatics, and imaging informatics will benefit from an integrated approach of piecing together different aspects of personalized information from a diverse range of data sources, both structured and unstructured, covering genomics, proteomics, metabolomics, as well as imaging, clinical diagnosis, and long-term continuous physiological sensing of an individual. It is expected that recent advances in big data will expand our knowledge for testing new hypotheses about disease management from diagnosis to prevention to personalized treatment. The rise of big data, however, also raises challenges in terms of privacy, security, data ownership, data stewardship, and governance. This paper discusses some of the existing activities and future opportunities related to big data for health, outlining some of the key underlying issues that need to be tackled.

Boshra Bahrami, Mirsaeid Hosseini Shirvani et al. Authors have evaluated different classification techniques in heart disease diagnosis. Classifiers like J48 Decision Tree, K Nearest Neighbors (KNN), Naive Bayes(NB), and SMO are used to classify dataset. After classification, some performance evaluation measures like accuracy, precision, sensitivity, specificity, F-measure and area under ROC curve are evaluated and compared. The comparison results show that J48 Decision tree is the best classifier for heart disease diagnosis on the existing dataset.

Umair Shafique, Fiaz Majeed Et al. Data mining in Healthcare is a crucial and complicated task that needs to be executed accurately. It attempts to solve real-world health problems in diagnosis and treatment of diseases. This work is also an attempt to find out interesting patterns from data of heart patients. There is three algorithms used with two different scenarios. These implemented algorithms are Decision Tree, Neural Network, and Naïve Bayes.

3. PROBLEM FORMULATION

Prediction analysis involves extracting information from a given dataset and using it to predict future results and further performance patterns. This technique is usually applied to past or present data whereas the prediction result is in future. For Instance, recognizing traces after any damage has been committed. However, the main key factor is, that closeness of results is directly proportional to the principal of analysis and accuracy of supposal. Evaluating at a more granular level can be determined as Predictive analytics. There has always been a curiosity to predict results of Matches and lot of Research work have been performed in Data Science to determine Models and logics that gives results of Heart Disease basis the Permutation and Combination of essential Parameters. These Models can be helpful in strengthening the health care industry enhance the prospects of early detection of disease. Gradually, with further advancements, it was observed that solely considering these factors does not solve the purpose of Prediction Analysis. Our study devised a model, which analyzes the heart disease in given dataset. It primarily does a feature selection and identifies most significant features of the health. Creation and Implementation of Fuzzy Rules to calculate their Impact on Result follow this. Thereafter, we divide this Data into Training and Testing Data Sets and employ Naïve Bayes Technique to obtain prediction accuracy. In contrast with last studies in determining the outcomes of heart disease, the prediction closeness retrieved from our model is marginally large.

4. PROPOSED SYSTEM

The two most normal modeling goals of data mining are classification and prediction. Classification models classify discrete, unordered values or data whereas prediction models predict continuous valued. Decision trees and Neural Networks are examples of classification models while Regression, Association Rules, and Clustering are examples of the prediction algorithm. In the proposed system of heart disease, we will use the following algorithms of data mining are analyzed:

A. GA

B. Naïve Bayes

In proposed system we would use data mining techniques brings with a set of techniques to find out hidden patterns for making Predictions in Heart Disease. The proposed system will be focused on classification & feature selection methods of data mining used in data discovery. Feature selection techniques would be used to extract the significant features. Classification techniques of data mining would be used for data classification and knowledge extraction. The proposed system would predict the heart disease based on different attributes of Patient's lifestyle, diet, family history etc. The system analyzes the pattern, which would be made by inputting data. After analyzing the system would give the prediction for a particular patient.

5. OBJECTIVES

- To study the literature survey for better understanding and to find out the drawbacks of previous systems.
- To study the classification and clustering models.
- To build the system, would be capable of pattern recognition of data.
- To train the system by inputting the data of patient's life-style/habits.
- Obtain the results by testing the data.
- Calculate the system accuracy by testing on a large dataset.

6. REFERENCES

- [1] Gandhi, Monika, and Shailendra Narayan Singh. "Predictions in heart disease using techniques of data mining." *Futuristic Trends in Computational Analysis and Knowledge Management (ABLAZE), 2015 International Conference on*. IEEE, 2015.
- [2] Abdar, Moloud, et al. "Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases." *International Journal of Electrical and Computer Engineering (IJECE)* 5.6 (2015): 1569-1576.
- [3] Purusothaman, G., and P. Krishnakumari. "A survey of data mining techniques on risk prediction: Heart disease." *Indian Journal of Science and Technology* 8.12 (2015).
- [4] Asmi, Shabana P., and S. Justin Samuel. "An analysis and accuracy prediction of heart disease with association rule and other data mining techniques." *Journal of Theoretical and Applied Information Technology* 79.2 (2015): 254.
- [5] Bandyopadhyay, Sunayan, et al. "Data mining for censored time-to-event data: a Bayesian network model for predicting cardiovascular risk from electronic health record data." *Data Mining and Knowledge Discovery* 29.4 (2015): 1033-1069.
- [6] Andreu-Perez, Javier, et al. "Big data for health." *IEEE Journal of biomedical and health informatics* 19.4 (2015): 1193-1208.
- [7] Bahrami, Boshra, and Mirsaeid Hosseini Shirvani. "Prediction and Diagnosis of Heart Disease by Data Mining Techniques." *Journal of Multidisciplinary Engineering Science and Technology (JMEST)* 2.2 (2015): 164-168.
- [8] Makwana, Ankur, and Jaymin Patel. "Decision Support System for Heart Disease Prediction using Data Mining Techniques." *International Journal of Computer Applications* 117.22 (2015).
- [9] Shafique, Umair, et al. "Data mining in healthcare for heart diseases." *International Journal of Innovation and Applied Studies* 10.4 (2015): 1312.
- [10] Assaleh, K., "Extraction of fetal electrocardiogram using adaptive neuro-fuzzy inference systems", *Trans. on Biomed. Eng.*, vol. 54(1), pp. 59-67, IEEE 2007.
- [11] Camps, G., Martinez, M., & Soria, E., "Fetal Ecg extraction using an FIR neural network", *Computers in Cardiology*, pp. 249-252, IEEE 2004.
- [12] Camps, G., Martinez, M., Soria, E., Magdalena, R., & CalpeGuerrero, J., "Foetal ECG recovery using dynamic neural networks", *Art. Intel. Med.*, 31(3), pp. 197-209, IEEE 2004.
- [13] Mehmet Korurek, Berat Dogan, "ECG beat classification using particle swarm optimization and radial basisfunction neural network", *Expert Systems with Applications*, vol. 37, pp. 7563-7569, Science Direct, 2010.
- [14] Tsung-Ching Wu, King-Chu Hung, Je-Hung Liu, Tung-Kuan Liu, "Wavelet-based ECG data compression optimization with genetic algorithm", vol. 6, pp. 746-753, *J. Biomedical Science and Engineering*, 2013.
- [15] Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte, A data mining approach for prediction of heart disease using neural networks, *international journal of computer engineering and technology*, 2012.
- [16] Mohammad Taha Khan, Dr. Shamimul Qamar and Laurent F. Massin, A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining, *International Journal of Applied Engineering Research*, 2012.
- [17] N. Aditya Sundar, P. Pushpa Latha, M. Rama Chandra, performance analysis of classification data mining techniques over heart diseases data base, *international journal of engineering science and advanced technology*, 2012.
- [18] Dustin Y. Harvey and Michael D. Todd. "Automated feature design for numeric sequence classification by genetic programming." *Evolutionary Computation, IEEE Transactions*, 19(4): (2015): 474-489.
- [19] Fedotov, Aleksandra A., Anna S. Akulova, and Sergey A. Akulov. "Uncertainties in measurement of nonlinear dynamics in heart rate variability." In *6th European Conference of the International Federation for Medical and Biological Engineering*, pp. 102-105. Springer International Publishing, 2015.
- [20] Nada Lavrac, *Selected techniques for data mining in medicine*, Elsevier, 1999.
- [21] Shantakumar B. Patil, Y. S. Kumaraswamy, "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network"; *European Journal of Scientific Research*, ISSN 1450-216X Vol.31 No.4, 2009.