Heart Disease Detection Using Machine Learning Algorithm

Sayali Deshmukh[1], Sanhita Divekar[2], Atharv Joshi[3], Nilam Panmand[4], Prof.Aditi Das[5]

Nutan Maharashtra Institute Of Engineering and Technology[1-5]

Abstract: Today's modern world cardiovascular disease is the most lethal one. This disease attacks a person so instantly that it hardly gets any time to get treated with. So diagnosing patients correctly on timely basis is the most challenging task for the medical fraternity. A wrong diagnosis by the hospital leads to earn a bad name and loosing reputation. At the same time treatment of the said disease is quite high and not affordable by most of the patients particularly in India. The purpose of this paper is to develop a cost effective treatment using data mining technologies for facilitating data base decision support system. Almost all the hospitals use some hospital management system to manage healthcare in patients. Unfortunately most of the systems rarely use the huge clinical data where vital information is hidden. As these systems create huge amount of data in varied forms but this data is seldom visited and remain untapped. So, in this direction lots of efforts are required to make intelligent decisions. The diagnosis of this disease using different features or symptoms is a complex activity. In this paper using varied data mining technologies an attempt is made to assist in the diagnosis of the disease in question. Keyword : cardiovascular disease, data mining, intelligent decisions, symptom.

Keywords: Machine learning algorithm, Knn Classification, Heart disease prediction, Report generation.

I. INTRODUCTION

Today, many hospitals manage healthcare data using healthcare information system; as the system contains huge amount of data, used to extract hidden information for making intelligent medical diagnosis. The value of machine learning in healthcare is its ability to process huge datasets beyond the scope of human capability, and then reliably convert analysis of that data into clinical insights that aid physicians in planning and providing care, ultimately leading to better outcomes, lower costs of care, The main objectives of this research is to build Intelligent Heart Disease Prediction System that gives diagnosis of heart disease using historical heart database. To develop this system, medical terms such as sex, blood pressure, and cholesterol like 13 input attributes are used. 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. The data mining classification techniques viz. Neural Networks, Decision Trees.KNN Classification, and Naive Bayes are used.The healthcare industry collects huge amounts of healthcare data which, unfortunately, are not "mined" to discover hidden information for effective decision making. Discovery of hidden patterns and relationships often goes unexploited. Advanced data mining techniques can help remedy this situation. This research has developed a prototype Intelligent Heart Disease Prediction System (IHDPS) using data mining techniques, namely, Decision Trees, Naïve Bayes and Neural Network. Results show that each technique has its unique strength in realizing the objectives of the defined mining goals. IHDPS can answer complex "what if" queries which traditional decision support systems cannot. Using medical profiles such as age, sex, blood pressure and blood sugar it can predict the likelihood of patients getting a heart disease. It enables significant knowledge, e.g. patterns, relationships between medical factors related to heart disease, to be established. IHDPS is -based, user-friendly, scalable, reliable and expandable. It is implemented on the Java-Python platform by using KNN Classification Algorithm.

II.LITERATURE SURVEY

Very few systems use the available clinical data for prediction purposes and even if they do, they are restricted by the large number of association rules that apply. Diagnosis of the condition solely depends upon the Doctors' intuition and patient's records. Detection is not possible at an earlier stages. In the existing system, practical use of various collected data is time consuming. There are only few decision support systems available in medical industry whose functionalities are very limited. As mentioned earlier, medical decisions are made with doctor's intuition and not from the rich data from the medical database. Wrong treatment due to misdiagnosis causes serious threat in medical field. In order to solve these issues data mining solution was with help of medical databases was introduced. The leading cause for mortality and morbidity is cardiovascular disease [1]. Ahmed M. Alaa proposed machine learning techniques for Cardiovascular disease risk prediction. But they achieved maximum accuracy of 77%. As the dataset is unbalanced, there is a need to apply sampling techniques. But they directly applied Machine learning models on the dataset. [2]Stephen F. Weng[3] et.al studied application of machine learning algorithms to improve cardiovascular risk prediction. They shown that Machine-learning algorithms are successful in improving accuracy of cardiovascular risk prediction, but the required number of patient records must be more to achieve better results. [3]Rine Nakanishi

evaluated ML methods for improving the prediction rate of coronary heart disease (CHD). They applied machine learning approaches on 6814 patient records and achieved good accuracy rate. Senthilkumar Mohan[6] proposed a machine learning model that finds significant features for improving the prediction rate of cardiovascular disease. They tried with various combinations of features and achieved an accuracy of 88.7% with hybrid KNN Classification. [4]Mohammed Abdul Khaleel has given paper in the Survey of Techniques for mining of data on Medical Data for Finding Fre-quent Diseases locally. This paper focus on dissect information mining procedures which are required for medicinal information mining particularly to find locally visit illnesses, for example, heart infirmities, lung malignancy, bosom disease et cetera. In-formation mining is the way toward extricating information for finding inactive examples which Vembandasamy et al. performed a work, to analyze and detect heart disease. In this the algorithm used was Naive Bayes algorithm. In Naïve Bayes algorithm they used Bayes theorem. Hence Naive Bayes has a very power to make assumption independently. The used data-set is obtained from a diabetic research institutes of Chennai, Tamilnadu which is leading institute. There are more than 500 patients in the dataset. The tool used is Weka and classification is executed by using 70% of Percentage Split. The accuracy offered by Naive Bayes is 86.419%. [5]Costas Sideris, Nabil Alshurafa, Haik Kalantarian and Mo-hammad Pourhomayoun have given a paper named Remote Health Monitoring Outcome Success prediction using First Month and Baseline Intervention Data. RHS systems are effective in sav-ing costs and reducing illness. In this paper, they portray an up graded RHM frameworks, Wanda- CVD that is cell phone based and intended to give remote instructing and social help to mem-bers. CVD counteractive action measures are perceived as a basic focus by social insurance associations around the world. [6] Sathish Kumar and A. Padmapriya has given a paper named Prediction for similarities of disease by using ID3 algorithm in television and mobile phone. This paper gives a programmed and concealed way to deal with recognize designs that are covered up of coronary illness. The given framework utilize information min-ing methods, for example, ID3 algorithm. This proposed method helps the people not only to know about the diseases but it can also help's to reduce the death rate and count of disease affected peoples.

III.PROPOSED METHODOLOGY

A. PROPOSED SYSTEM

The Heart Disease Prediction application is an end user support and consultation project. Here, we propose a application that allows users to get instant guidance on their heart disease through an intelligent system. The application is fed with various details and the heart disease associated with those details. The application allows user to share their heart related issues. It then processes user specific details to check for various illness that could be associated with it. Here we use some intelligent data mining techniques to guess the most accurate illness that could be associated with patient's details. Based on result, the can contact doctor accordingly for further treatment. In general, the more trees in the forest the more robust the forest looks like. In the same way in the KNN Classification classifier, the higher the number of trees in the forest gives the high accuracy result.
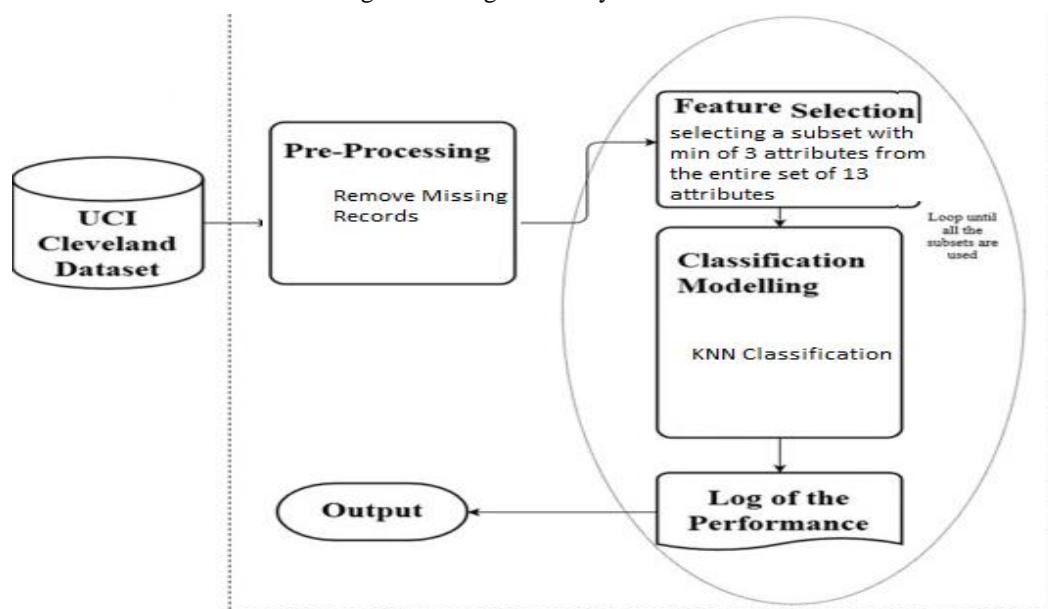
System Architecture

IV. CONCLUSION

Proposed algorithm, is a algorithm for predicting the risk of heart attack in a patient using the attributes collected from the data sets. Proper adaptation of the data into an optimum number of clusters and aids with detecting the normal and abnormal cases efficiently.Identifying the processing of raw healthcare data of heart information will help in the long term saving of human lives and early detection of abnormalities in heart condition.

V. ACKNOWLEDGMENT

REFERENCES

[1] J Thomas MR, Lip GY. Novel risk markers and risk assessments for cardiovascular disease. Circulation research. 2017; 120(1):133–149. https://doi.org/10.1161/CIRCRESAHA.116.309955 PMID: 28057790

[2] Ahmed M. AlaaID1, Thomas Bolton, Emanuele Di Angelantonio, James H.F. RuddID, Mihaela van der Schaar,―Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants‖, PLOS ONE 14(5): e0213653. https://doi.org/10.1371/journal, May 15, 2019H. Poor, ―A Hypertext History of Multiuser Dimensions,‖ MUD History, http://www.ccs.neu.edu/home/pb/mudhistory.html. 1986. [3] Stephen F. Weng, Jenna Reps, Joe Kai1, Jonathan M. Garibaldi, Nadeem Qureshi, ―Can machine-learning improve cardiovascular risk prediction using routine clinical data?‖, PLOS ONE | https://doi.org/10.1371/journal.pone. [4] Rine Nakanishi, Damini Dey, Frederic Commandeur, Piotr Slomka, ―Machine Learning in Predicting Coronary Heart Disease and Cardiovascular Disease Events: Results from The Multi-Ethnic Study of Atherosclerosis (Mesa)‖, JACC Mar- 20, 2018, Volume 71, Issue 11

[5] https://www.cdc.gov/heartdisease/facts.htm. Available [Online].

[6] Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava ―Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques‖, Digital Object Identifier 10.1109/ACCESS.2019.2923707, IEEE Access, VOLUME 7, 2019 S.P. Bingulac, ―On the Compatibility of Adaptive Controllers,‖ Proc. Fourth Ann. Allerton Conf. Circuits and Systems Theory, pp. 8-16, 1994. (Conference proceedings).