iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** III **Month of publication:** March 2020

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue III Mar 2020- Available at www.ijraset.com

Prediction of Heart Disease using Data Mining

S. M. Bhadkumbhe[1], Vaishnavi Metkar[2], Kshitija Shitole[3], Harshwardhan Suryawanshi[4], Ruchita Prabhune[5]
[1, 2, 3, 4, 5]*Computer Department, PDEA's College of Engineering Pune*

Abstract: ***Data mining process involves mining of very significant, hidden and valuable information from large datasets. Usually the Healthcare sector involves huge amount of data related to patients, various diagnosis of the diseases etc. Nowadays, culture of hospital is to use information management systems in order to handle patient's data systematically and effectively. Large quantity of data is produced by such systems that is represented using charts, numbers, text and images. Though such sort of data is hardly employed for making any clinical decisions. The current research emphasizes on heart disease diagnosis. Various techniques of data mining have been incorporated for diagnosing the disease thereby obtaining several probabilities. The main aim of proposed system is to verify the doctor's result. The system consider some attributes for detection of disease. Our project is perspective to design and develop futuristic prediction regarding details of health system for heart diseases based on predictive mining. There are various experiments have been conducted to a specific relationship for the performance of various predictive data mining techniques including Decision tree, Naïve Bayes and k-mean algorithms. In this proposed work, attribute structured clinical database has been used as a source data. Naive Bayes have been prescribed and their schedule on diagnosis has been compared.***

Keywords: Data Mining, Naïve Bayes, k-mean, Decision tree

I. INTRODUCTION

One of the important part of body is heart which supply blood to other parts of body for their proper functioning. Heart co-ordinate with other part of body for blood circulation. If the proper blood is not supplied to other parts of body then it leads to fatal issue. So the proper working of heart is necessary. The symptoms of any disease vary from person to person. In the case of heart disease the symptoms vary from person to person depending upon the type of heart disease. Coronary heart disease is the most common type of heart disease. It is occurred due to lack of blood supply to the heart. Coronary heart disease is the leading cause of death. Tremendous amount of people are prone to heart deformity and in today's world lot of people are prone to these diseases that is Cardiovascular disease and it is one of the major reasons for heart problem. Data mining is a great developing technique that revolves around exploring and digging out significant information from massive collection of data which can be further beneficial in examining and drawing out patterns for making decisions. Talking about the Medical domain, implementation of data mining in this field can yield in discovering and withdrawing valuable patterns and information which can prove beneficial in performing clinical diagnosis. This research focuses on heart disease diagnosis by considering previous data and information. The speedy advancement of technology has led to remarkable rise in mobile health technology that being one of the web application. The required data is assembled in a standardized form. For predicting the chances of heart disease in a patient, complex clinical data is driven by technologies such as health care about patients and other hospital resources. Rich collection are used in order to Data mining techniques examine methodology in detail the different perspectives and deriving useful detail.

Concerning the heart disease prediction numerous systems are being recommended which are being deployed by the means of various techniques and algorithms. Gaining quality service at affordable price remains the prime and challenging concern for the healthcare establishments. For offering quality services at par, there must be accurate diagnosis of the patients along with effective dosage of medicines. Low quality clinical diagnosis and treatment can yield in undesired and inadequate results. One solution for cost cutting by Healthcare establishments can be utilization of computer generated data or use of decision support systems. Usually the Healthcare sector involves abundant of data related to patients, various diagnosis of the diseases, resource management etc. This information or data must be further broken down by the Human services. Using computerized system, patients treatments records can be stored and using mining methods one can acquire significant information and queries concerning the hospital. Supervised and unsupervised learning are the two data mining methods. Supervised learning involves usage of training for learning model parameters where else no training set is required in unsupervised learning. Classification and prediction are the basic approach of data mining. The Classification models helps in classifying distinct, disorganized data values on the other hand prediction model anticipated values that are continuous. Thereafter making use of the analysis result for offering web/mobile application to the users.

In today's world the hospitals manages the patient's data on the system. All of this data is stored in huge databases of electronic medical records systems.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue III Mar 2020- Available at www.ijraset.com

This system generate huge amount of data in different formats. But this data are not used for further processing of clinical decision. So the main aim is to utilize the clinical data for prediction of disease. Data mining for health care utilizes the medical records for the prediction of heart disease specially for CVD.

II. PROPOSED WORK

As per today's advanced and hi-tech living style, majority of the people are contracting heart disease which gives a sudden jolt to an individual that at times one lacks time to get treated immediately. Hence its very much essential that timely and early diagnosis is performed which being quiet challenging concern for the medical association. Poor and incorrect analysis carried out by the hospital can being down its reputation and working. The research focuses on to build cost cutting and effective approach by the means of data mining techniques so that decision support system can be enhanced. Predicting heart disease with the help of numerous attributes/symptoms is quiet complicated. The present research utilizes Naive Bayesian - data mining classification technique for effectively enabling heart disease diagnosis and thereby offering appropriate treatment. Supervising different medical factors and post operational period stands very crucial. Patients' records are save in database. The results generated reveals that the diagnostic system built successfully predict the risk level associated with heart diseases.

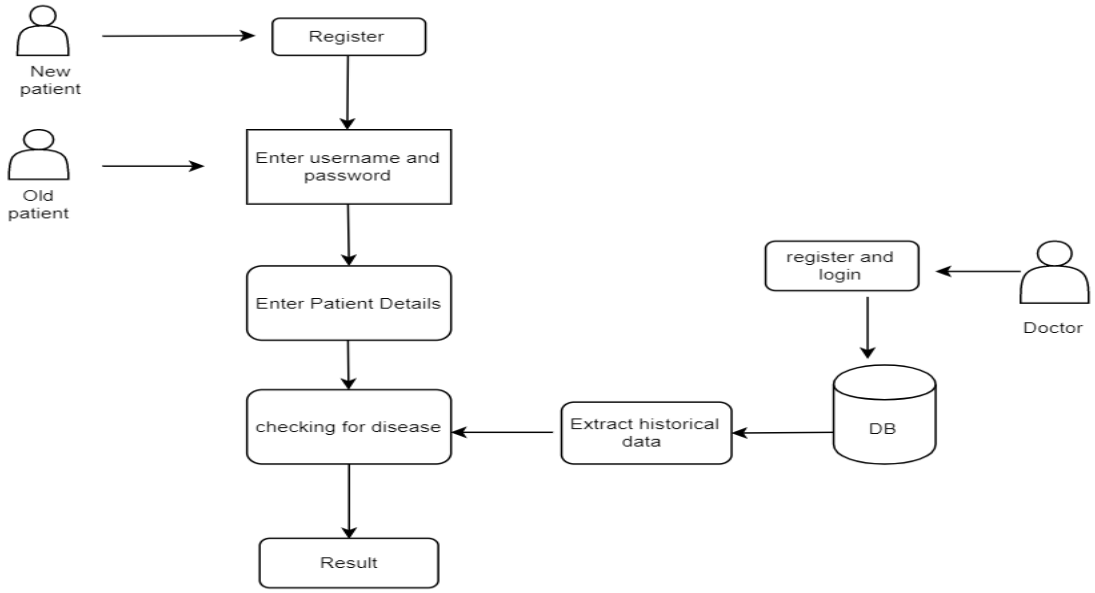


Fig. 1 Architecture view

A. Patient

The most important user of system is patient. Here patient can check for presence or absence of disease. The patient first register himself/herself to access the system. The registration form for patient consist of patient id, name, age, gender, email, contact number. Patient can update this information as and when required. Also we can delete the patient record. After registration patient should enter his/her id and login to the system. Already register user can access system by entering id and password. After login patient enter the factors values so that the result can be calculated. The values should be in the mention criteria. It consist of different risk factor like age, chest pain, gender, rating blood pressure, cholesterol, ECG ect.

B. Doctor

In today's world doctors plays a very important role for detection as well as for diagnosis of that detected disease. But at the same time the fraud in the medical sector is increasing so we have to take care of it. In the proposed system the doctor register themselves by entering some details like id, name, speciality etc. Then this information is used by patient for diagnosis of detected heart disease.

C. Dataset Collection

Using the UCI dataset, Collection of medical data of patients with heart diseases is carried out. Every patient's attributes are being assembled such as demographic, historic and laboratory features such as sex, age, hypertension, diabetes mellitus, chest pain type, random blood sugar, cholesterol, systolic and diastolic blood pressure, old peak, thal ect. are consider as the training dataset.

D. Heart Disease Prediction Factor

Different types of attributes have been used for better prediction of heart disease. The following are the factors consider for prediction of heart disease. They are classified into Common Factors and Medical Oriented Factors for better understanding. They are as follows:

1) Common Factor: Common Factors are the factors in which the values cannot be changed and obtained without any medical test.

a) Age in Years: Heart disease may occur at any phase of life. Specifically, in case of men heart attack might occur after the age of 45 and for women it occurs after the age of 55. The people in 20's and 30's also suffer from heart attack. (values: Age in years (AGE)-Young (YNG) <33, Medium (MED) 34–40, Old (OLD) 41–52, Very Old (VOLD) >52)

b) Gender: It describes the gender (value 0 or 1) 1: Male; 0: Female

c) Chest Pain Type: Angina is a pain in the chest, caused when oxygen rich blood is not properly supplied to the heart muscle. Non-Angina chest pain is likely a chest pain with duration over 30 minutes or less than 5 seconds. (values: Typical Angina (TA) -1, Atypical (ATA) -2, Non-Angina pain (NAP)- 3, Asymptomatic (ASY) -4)

d) Induced angina: Induced angina occurs when the heart muscle does not get the oxygen it needs to function properly when you exercise. (values: True (YES) 1, False (NO) 0)

2) Medical Factor: Medical Oriented factors are the factors in which the values can be changed and obtained from the medical test results.

a) Fasting Blood Sugar: A carbohydrate metabolism test is used to measure the blood sugar level. It is conducted after fasting. Where normal is 70 -108 mg/dl; High blood glucose level is 109 and above. (value 0 or 1)

b) Resting electrocardiographic (ECG): The Resting electrocardiographic test measures the heart's electrical activity. Normally, the electric impulses cross the heart which contracts approximately 60 - 80 times per minute at rest. The left ventricular hypertrophy can be diagnosed when there is a large mass of myocardium for electrical activation to pass through (values: normal-0, ST-T wave abnormality1, definite left ventricular hypertrophy-2)

c) Slope: The Isoelectric section of the ECG which represents the S-wave, T-wave and the ST segment (values: Up-sloping (UPS) 1 Flat (FLT) 2 Downs-loping (DWS) 3)

d) Major vessels: It represents the count of the dominant vessels coloured by fluoroscopy (values: Fluoroscopy-0 (FL-0) 0 Fluoroscopy-1 (FL-1) 1 Fluoroscopy-2 (FL-2) 2 Fluoroscopy-3 (FL-3) 3)

e) Thal: Thallium Scan taken when the heart's ability to pump is impaired. In a normal flow, blood is supplied to the entire heart muscle. In reversible defect the blood flows less to the heart muscle which takes place due to insufficient blood supply from a specific coronary artery. In fixed defect, the blood flows less to the heart muscle which takes place as a result of permanently damaged muscle (Values: Normal (NOR) 3 Fixed defect (FDE) 6 Reversible defect(RDE) 7)

f) Cholesterol (mg/dl): The Serum Cholesterol(SC) contains triglycerides, high density lipoprotein and low density lipoprotein. The normal level of Serum Cholesterol is 200 mg/dl

g) Heart rate: A valuation of a person's highest age-related value can be obtained from deducting the age of the person from 220. (values: Low (LOW) <112 Medium (MED) 112–152 High (HIGH) >152)

h) Old Peak: ST depression which is caused due to exercise associated to rest. Coronary insufficiencies are the major cause for ST depression. While an exercise stress test requires an ST depression of at-least 2mm to significantly indicate reversible inchaemia.(values: Low (LOW) <1.5 Risk (RSK) 1.5–2.55 Terrible (TER) >2.55)

i) Resting Blood Pressure: It is the pressure as the heart relaxes. The normal blood pressure is defined as 120mm systolic and 80mm diastolic. (values: Low (LOW) <128 Medium (MED) 128–142 High (HIGH) 143–154 Very High (VHIGH) >154)

E. Calculation of Result

After getting the input risk factors from patient the datamining phases are performed on the training dataset. Then on training data algorithms such as decision tree, k-means and naïve Bayesian are applied to get the results. This classification algorithm basically employs conditional independence, this implies that value of an attribute for an available class is not dependent on other attribute values since the algorithm relies upon the Bayesian theorem. The classifier is based on conditional independence, this implies that value of a variable for an available class is independent of other existing variable value. In case of high dimensionality input, the classifier is highly appropriate. Using Naïve Bayesian, models having predictive potentials can be designed. The k-means algorithm used for clustering of similar types of data. K-means algorithm is mainly used to find the values of the factors of heart disease by using the k user defined value. By using the K values, it is possible to make boundaries for each class of heart disease related attributes. Decision tree is a type of supervised learning algorithm. It works for both categorical and continues input and output

variables. In the diagnosis of heart disease, decision tree will segregate the datasets based on all the values of the attributes and identify the attribute, which creates the best homogeneous sets of data. It also provides classified report for the heart disease. After calculating the probability of each class attribute by naïve Bayesian algorithm the result is displayed to the user.

III.FUTURE WORK

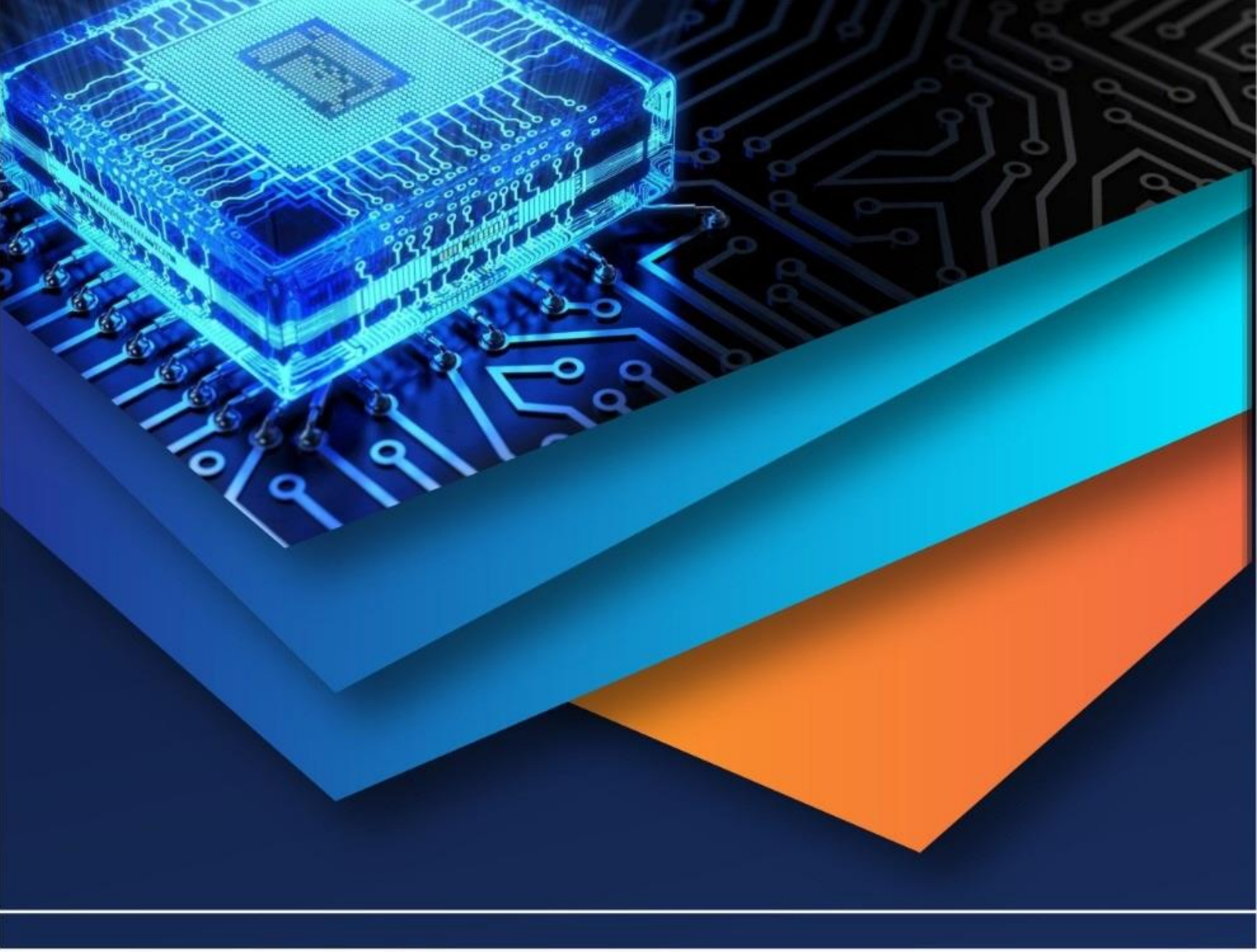
This research work mainly covers prediction of heart disease using UCI dataset, which can be further enhanced using real time dataset. We can add the security encryption and decryption protocols for user's data. By implementing the big data concepts all the data form different medicals stored on same cloud so can anyone access it as the historic data. Also we can consider more attributes to get the highly accurate results with different combination of algorithms.

IV.CONCLUSIONS

Naïve Bayesian Classification technique is used to develop decision support in Heart Disease Prediction System. The UCI dataset are used for prediction of disease. There are several attributes in the dataset but we consider only dominant attribute which is mostly responsible of heart disease. The algorithms used for prediction have high accuracy. Anyone can use this system because it is easy to understand. Using this system medical students and nurses can be trained to diagnose heart disease. This system can help doctors to make better decisions and can be used to provide a second opinion. The hidden knowledge can be extracted from a heart disease database. The result should be extensively tested by experienced cardiologists. By implementing this system we can find fraud in medical sector also it is cost efficient.

REFERENCES

[1] Sudeshna, P., S. Bhanumathi, and MR Anish Hamlin. "Identifying symptoms and treatment for heart disease from biomedical literature using text data mining." Computation of Power, Energy Information and Commuincation (ICCPEIC), 2017 International Conference on. IEEE, 2017.
[2] Sonet, KM Mehedi Hasan, "Analyzing patterns of numerously occurring heart diseases using association rule mining." Digital Information Management (ICDIM), 2017 Twelfth International Conference on. IEEE, 2017
[3] Sowmiya, C., and P. Sumitra. "Analytical study of heart disease diagnosis using classification techniques." Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2017 IEEE International Conference. IEEE, 2017R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.
[4] Babu, Sarath, "Heart disease diagnosis using data mining technique." Electronics Communication and Aerospace Technology (ICECA), 2017 International conference of .Vol. 1. IEEE, 2017
[5] Beant Kaur h, Williamjeet Singh, "Review on Heart Disease Prediction System using Data Mining Techniques", © IJRITCC, Vol.2, Issue: 10, p.p.3003-08, 2014
[6] Ashok Kumar Dwivedi "Performance evaluation of different machine learning techniques for prediction of heart disease", © Springer, Computer Applications and Mathematics, 2016.
[7] Chala Beyene, Pooja Kamat (2018), "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, Vol. 118, No. 8, pp. 165-174
[8] Kamal Kant, and Dr. Kanwal Garg, 2014, "Review of Heart Disease Prediction using Data Mining Classifications", © IJSRD, Vol. 2, Issue 04, pp. 109-111
[9] Mrs. G. Subbalakshmi,Mr. K. Ramesh and Mr. M. Chinna Rao, Decision Support in Heart Disease Prediction System using Naive Bayes,IJCSE1102-02-56
[10] Ankita R. Mokashi1, Madhuri N. Tambe2, Pooja T. Walke, Heart Disease Prediction Using ANN and Improved K-Means,IJIREEICE Vol. 4, Issue 4, April 2016



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY