IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-12, Dec 2018 E-ISSN: 2347-2693

Prediction of Heart Disease Using Data Mining Technique

[1*]Ranjana Joshi, [2]Shilpa Sethi

[1]Research scholar, Department of Computer Engineering, YMCA University of Science &Technology, Faridabad
[2]Assistant Prof, Department of Computer Application, YMCA University of Science &Technology, Faridabad

*Corresponding Author: ail: 12joshi.ranjan@gmail.com

Available online at: www.ijcseonline.org

Accepted: 16/Dec/2018, Published: 31/Dec/2018

***Abstract*-** In the past few decades, heart diseases had been found as leading cause of death across the world. But at the same time, it is also discovered as most preventable and controllable disease. Further, it is identified that its early and timely diagnosis helps in controlling its growth and reducing treatment cost remarkably. So it become essential to discover accurate and reasonable tools capable of extracting high risk data for timely diagnosis of such a critical disease The proposed work aims to identify vital parameters leading to heart diseases and develop a model based on data mining techniques.

***Keywords*-** Data mining, Heart disease Prediction, Naive Bayes, Technique.

I. INTRODUCTION

Extraction of useful information from huge amount of data is known as data mining [7]. Data mining is the process of finding previously unknown patterns and trends in databases and using that information to build predictive models [8]. Nowadays to live luxurious life people work like machine in order to earn lot of money hence they forget to take care of their health. Because of this, there is change in the food which they consume, their lifestyle changes. They are more tensed under the pressure of earning more money. This leads to various diseases like blood pressure, diabetes, cholesterol at a very young age. All these reasons lead further to negligence of their health which increases the chances of heart disease. Heart is the most essential organ of the human body and if it gets affected then it also affects the other major organs of the body. The aim of proposed system is to design an automated system which would manage complete clinical details, patient's history and their appointment details in a single database. Doctors will use this system to keep track of the patient consulting to them. The intentions of the proposed system are to reduce over-time pay and increase the number of patients that can be treated accurately.

The paper is organized as follows: Section II covers the literature review in the area of clinical diagnostic system. The section also identifies the shortcomings of existing system. Section III covers the proposed system. Section IV covers implementation snapshots. Section V concludes the research work.

II. RELATED WORK

Many hospital information systems are designed to support patient billing, inventory management and generation of simple statistics. Some hospitals use decision support systems, but are largely limited. They can answer simple queries like:

"What is the average age of patients who have heart disease?", "How many surgeries had resulted in hospital stays longer than 10 days?", "Identify the female patients who are single, above 30 years old, and who have been treated for cancer." However they cannot answer complex queries like:

"Given patient records, predict the probability of patients getting a heart disease."

Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the database. It has been observed that problems like: (i) Unwanted biases, (ii) Medical errors (iii) Large medical costs which affect the quality of service provided to patients. The proposed system aim to resolve these problems by integrating decision support system and computer based potential record systems. The system based on concepts of data mining, has the potential to generate a knowledge rich environment which helps to significantly improve the quality of clinical decisions.

Quality service for clinical system implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are obviously unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems. 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 like decision tree and naïve bayes can help remedy this situation.

2.1 Decision Tree

Decision tree is a tree like structure that consists of nodes and edges. Upper most nodes is called Root node and sub-nodes are called children node [5]. Root has only one incoming edge. Nodes that do not contain outgoing edges are called terminal nodes. Various applications of decision tree include:

1. Sentiment Analyzers
2. Investment Solutions
3. Large number of Customer Satisfaction

2.2 Naïve Bayes

Naïve Bayes is a classification technique based on the Bayes theorem [1]. Bayes theorem is used to give probabilistic value for occurrence of an event. The various applications of Naïve Bayes include:

1. Face recognition
2. Sentiment analysis
3. Medical diagnosis
4. Digit recognition
5. Weather prediction

Based on the available literature, The proposed methodology uses Naïve Bays for heart disease prediction. The next section discusses the proposed method in detail.

III. PROPOSED SYSTEM

The heart disease prediction system based on data mining consists of four functional components as depicted in Fig 1. These are as follows:

1). Data collection module.
2). Data classifier.
3). Data mining module.
4). Medical database.

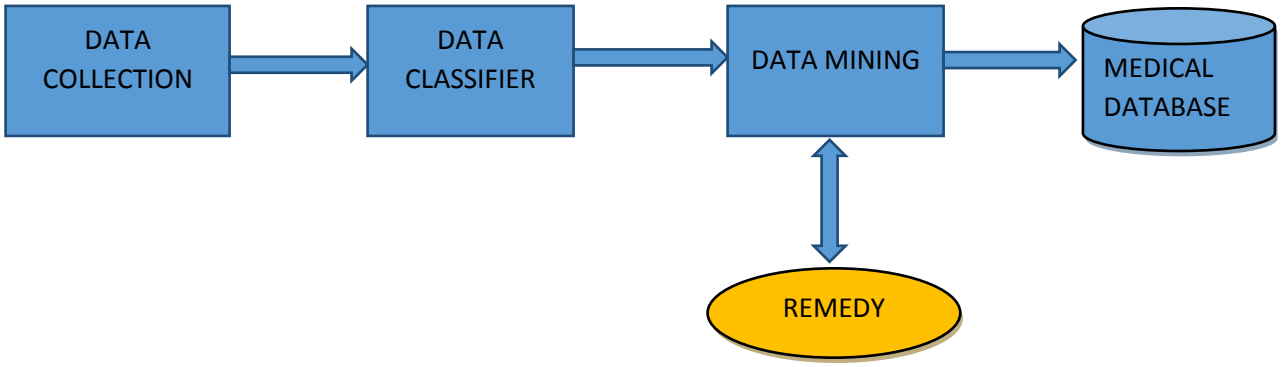

Fig 1 Architecture of Heart disease Prediction System using Data Mining

The detail working of each of these components is given in following subsections.

3.1 Data Collection Module

The major task of data collection module is to gather the patient data relevant to heart disease. Various methods can be adaptable for this purpose. Two most common are: -

1. Interviewing and
2. Patient history.

For interviewing, questions are prepared. For this purpose, the proposed system has used patient data of Cleveland clinic foundation [9]. It is freely available in UCI repository.

3.2 Data Classifier

The collection data is classified into set of certain attributes. The choice of attributes is purely based upon the factors that lead to heart disease. For this purpose, 13 different attributes are identified. These attributes are listed in table1.

Table 1: Various Parameters for Heart Disease Prediction

S No.	Risk factor	Values
1	**Sex**	**Male or Female**
2	Age in Years	Age in Numeric
3	Blood Cholesterol	Below 120 mm Hg-Low

		120 to 139 mm Hg-Normal Above 139 mm Hg-High
4	Blood Pressure	Below 120 mm Hg-Low 120 to 139 mm Hg-Normal Above 139 mm Hg-High
5	Hereditary	Family Member diagnosed with HD Yes Otherwise No
6	Smoking	Yes or No
7	Alcohol Intake	Yes or No
8	Physical Activity	Low , Normal or High
9	Diabetes	Yes or No
10	Diet	Poor , Normal or Good
11	Obesity	Yes or No
12	Stress	Yes or No
13	Medical Diagnosis	Heart Disease Yes or No

3.3 Data Mining Module

It is the most important component of proposed system. It is based on Naïve Bayes theorem. According to the Bayes theorem the probability of hypothesis H to be true for a given evidence E can be complete using the eq.1.

$$(H/E) = \frac{P(E/H)}{P(E)} * \mathrm{P(H)} \quad \ldots (1)$$

Where:-

- *P(H|E)* represents the probability of a hypothesis to be true for given evidence E.
- *P(H)* refers to prior probability of class.
- *P(E/H)* represents the inverse probability.
- *P(E)* refers to prior probability of predictor.

3.4 Medical Database

The data or information is stored and indexed in medical database. The schema for the same is given in Fig 2. The description of each field of schema is given in Table 2.

Identifier	Disease	Symptoms	Remedy

Fig 2: Schema for Medical Database

Table 2: Description of various Fields of Medical Database

Term	Definition
Identifier	It is an alphanumeric number that identifies a particular disease like d1, d2 and so on.
Disease	It symbolizes the particular disease like Cancer, Heart attack, Typhoid, Diabetes etc.
Symptoms	It represents Symptoms related to specific disease e.g. symptoms related to cancer are: Skin changes, Breast changes, Changes in bowel habits, Difficult or painful urination and symptoms related to heart attack are Chest discomfort, Discomfort in other areas of the upper body, hotness of breath, Other signs like nausea or light headedness.

IV. IMPLEMENTATION

The proposed heart disease prediction system using data mining is implemented using Python version 3.7.0. The patient's records of Cleveland clinic foundation available in UCI repository is used for analysis purpose. The various functional tabs used in detection of heart disease is depicted in implementation screenshot in Fig 3.

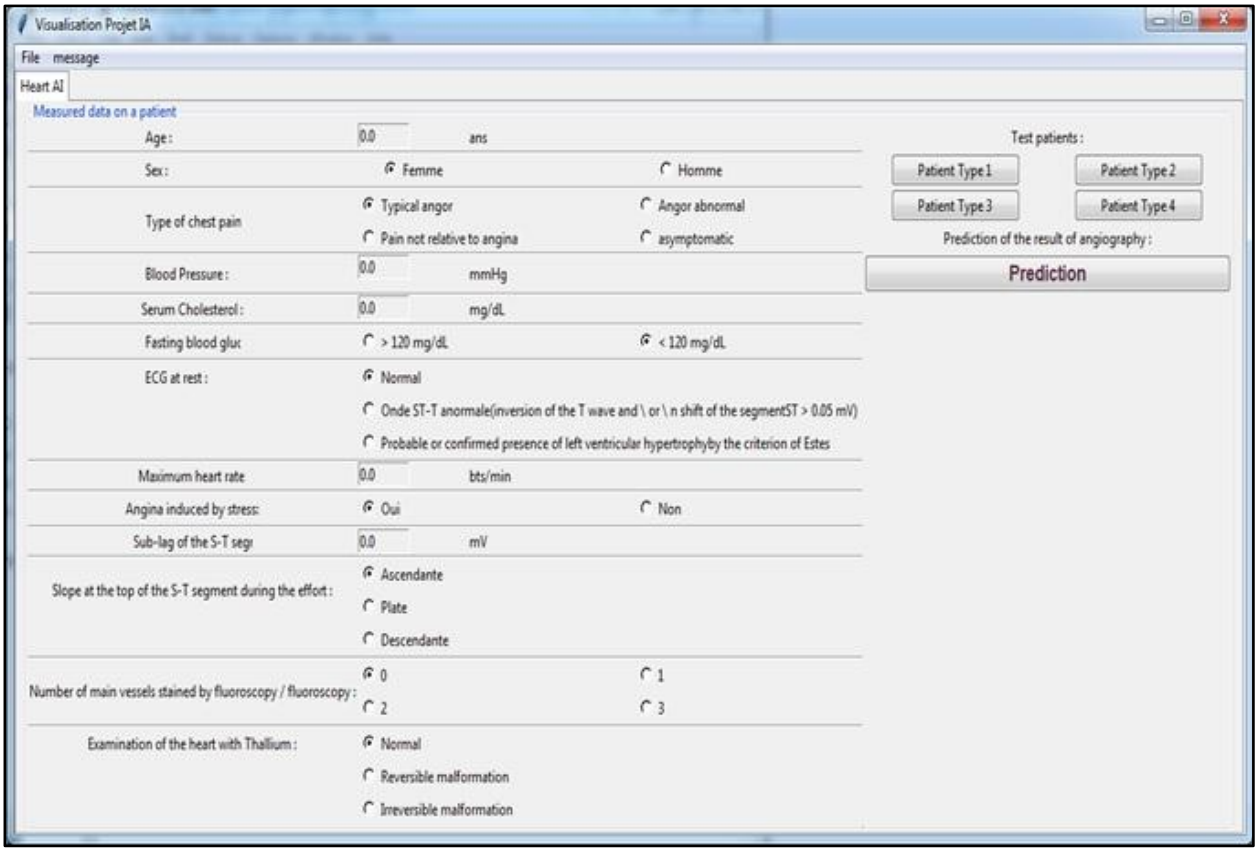

Fig 3: Implementation Screen shot of Proposed System

By applying the equation (1) on various parameters given in Table 2 and taking a threshold of 0.5 it is observed that system is able to predict chances of having heart disease fairly above average.

V. CONCLUSION

The paper presented an intelligent and effective heart disease prediction method using data mining. The system had used efficient approach for the extraction of significant patterns from the medical data warehouses for the prediction of heart disease. Medical diagnosis is considered as a significant yet intricate task that needs to be carried out precisely and efficiently. The automation of the same would be highly beneficial. The work shows that data mining has potential to generate useful knowledge from raw data of patients which further help to significantly improve the quality of clinical decisions.

REFERENCES

[1] Shadab Adam Pattekari and Asma Parveen, "Prediction System for Heart using Naive Bayes", *International Journal of Advanced Computer and Mathematical Sciences* ISSN 2230-9624.Vol 3, Issue 3, 2012, pp 290-294.

[2] 'http://www-igm.univ-mlv.fr/~lecroq/string/node14.html', Boyer Moore, accessed on 25th April,2014.

[3] SellappanPalaniappan, RafiahAwang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IJCSNS

International Journal of Computer Science and Network Security, Vol.8 No.8, August 2008.
[4] Anthony M. Dymond, "Data Extraction Methods in Medium and Large Databases", Dymond and Associates, LLC, Concord.
[5] Ansari Dipesh Sharma, SunitaSoni. "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction" International Journal of Engineering Research & Technology .
[6] Chaitrali S. DangareSulabha S. Apte, Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques" International Journal of Computer Applications (0975 – 888).
[7] S. Sethi, A. Dixit, "An Automatic User Interest Mining Technique for Retrieving Quality Data", *International Journal of Business Analytics*. Volume 4 • Issue 2 • April-June 2017, pp 62-79, ISSN: 2334-4547 .
[8] S. Sethi, A. Dixit, "An Adaptive Web Search System Based on Web Usage Mining", *International Journal of Computer Engineering and Applications (IJCEA*) Vol-X • Issue- I • Jan 2016, pp 09-18, ISSN: 2321-3469.
[9] https://archive.ics.uci.edu/ml/datasets/heart+Disease

[10] Priyanka, Sana Khan, Tulsi Kour, "Investigation on Smart Health Care Using Data Mining Methods", International Journal of Scientific Research in Computer Science and Engineering, Vol.4, Issue.2, pp.31-36, 2016.

Authors Profile

. Ms. Ranjana Joshi pursued Bachelor of Technology from Dev Bhoomi Institute of Technology affiliated from Uttrakhand Technical University, Dehradun in 2014 and she is currently pursuing Master of Technology (Information Technology) from YMCA University of Science & Technology,Faridabad in Department of Computer Engineering.

Dr. Shilpa Sethi has completed her PhD in 2018 from YMCA University of Science and Technology, Faridabad. She has received her Master in Computer Application from Kurukshetra University, Kurukshetra in the year 2005 and M. Tech. in Computer Engineering from MD University Rohtak in the year 2009. She has published more than 20 research papers in various International journals and conferences. Her area of research includes Internet Technologies, Web Mining and Information Retrieval System.

