



RESEARCH ARTICLE

HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES

Jyoti Rohilla¹, Preeti Gulia²

¹M.Tech Scholar, Dept of Computer Science & Applications, Maharshi Dayanand University, Rohtak

²Assistant Professor, Dept of Computer Science & Applications, Maharshi Dayanand University, Rohtak

¹rohillaajyoti93@gmail.com ; ²Preeti.gulia81@gmail.com

Abstract— Heart disease is a term that dispenses to an overprovision of anomalous health conditions that directly influence the heart and all its parts. Heart disease is the most important source of deaths widespread and the prediction of Heart Disease is significant at an inconvenient stage. As medical records systems turn out to be more consistent and ordinary, data quantity increases with much of it going unanalysed, which contains hidden information. This hidden information is useful for making effective decisions. Computer based information along with advanced Data mining techniques are used for appropriate results. This research paper explores the utility of Data mining techniques and to predict more accurately the presence of heart disease.

Keywords— Data mining, Cardiovascular disease, Classification.

I. INTRODUCTION

Cardiovascular disease is the most fatal one and highest-flying diseases of the modern world. Cardiovascular disease directly hits a person that it hardly gets any chance to get treated with. So diagnosing patients correctly on timely basis is the most challenging task for the medical fraternity.

Heart disease diagnosis is a multifaceted task which entails a great deal of experience and knowledge. A number of different tests are used to diagnose heart-related problems, including:

- Electrocardiogram (ECG)
- Exercise stress tests
- X-rays
- Echocardiogram
- Blood tests
- Coronary angiography
- Radionuclide tests
- Magnetic resonance imaging (MRI) scans

Some of tests are very expensive due to which treatment of the said disease is quite high and not affordable by most of the patients. However using data mining technique can reduce the number of test that is required. Data Mining is one of the most encouraging areas of research that turn out to be increasingly popular in health organization. Data Mining is significant for exposure of new drifts in healthcare organization which in turn useful for all the parties associated with this field. Data mining techniques play an important role in finding

patterns and extracting knowledge from large volume of data. It is very helpful to provide better patient care and effective diagnostic capabilities.

The risk factors for heart disease can be divided into modifiable and non modifiable.

- Modifiable risk factors include obesity, smoking, lack of physical activity and so on.
- Non modifiable risk factors for heart disease are like age, gender, and family history. Many people have at least one heart disease risk factor.

II. SIGNIFICANCE OF DATA MINING IN HEALTHCARE

Firstly, why data mining? Major problem is the availability of data. We are drowning in data but starving for knowledge. Data mining is a combined term for dozens of techniques to scrape together information from data and turn it into meaningful rules to improve perceptive of the data.

The vital ideology of data mining is to examine the data from different angle. Data mining tools predict future trends, allowing business to make proactive, knowledge-driven decisions. Data mining tools can answer questions that take too much time to resolve.

The huge amounts of data generated for prediction of heart disease are too intricate and voluminous to be processed and analyzed by traditional methods. Data mining provides the methodology and technology to transform these mounds of data into useful information for decision making. By using data mining techniques it takes less time for the prediction of the disease with more accuracy.

A number of algorithms have been developed and implemented to extract information and discover knowledge patterns that may be useful for decision support.

Data mining provides automatic pattern recognition and attempts to uncover patterns in data that are difficult to detect with traditional statistical methods. Data mining techniques form a group of heterogeneous tools and techniques and are used for different purposes. These techniques and methods are based on statistical techniques, visualization, machine learning, etc. People frequently use data mining to gain knowledge, not just predictions.

By applying data mining techniques in Healthcare domain, the administrators can improve the QoS (Quality of Service) by discovering latent potentially useful trends required by medical diagnosis.

A. *Data Mining Activities*

- Description and visualization;
- Association;
- Clustering;
- Classification and estimation

B. *Typical Healthcare Data Mining Applications*

- Treatment effectiveness;
- Healthcare management;
- Improving customer relationship management;
- Fraud and abuse detection;

C. *The Role of DM in Healthcare*

- Huge and complex volumes of data are generated by healthcare activities; un-automated analysis has become impractical;
- DM can generate information that can be useful to all stakeholders in health care, including patients by identifying effective treatments and best practices;
- The existence of insurance fraud and abuse impels insurers to use DM.

D. *Comparative Statement*

The following table presents the comparative statement of various data mining trends from past to the future taken from Beant Kaur et al [3].

TABLE I.
TABLE PRESENTS THE COMPARATIVE STATEMENT OF VARIOUS DATA MINING TRENDS FROM PAST TO THE FUTURE.

Data mining Trends	Algorithms/ Techniques Employed	Data formats	Computing Resources
Past	Statistical, Machine Learning Techniques	Numerical data and structured data stored in traditional databases	Evolution of 4G PL and various related techniques
Present	Statistical, Machine Learning, Artificial Intelligence, Pattern Reorganization Techniques	Heterogeneous data formats includes structured, semi structured and unstructured data	High speed networks, High end storage devices and Parallel, Distributed computing etc...
Future	Soft Computing techniques like Fuzzy logic, Neural Networks and Genetic Programming	Complex data objects includes high dimensional, high speed data streams, sequence, and noise in the time series, graph, Multi instance objects etc.	Multi-agent technologies and Cloud Computing

E. *Data mining used in Various Applications some of them are given below*

- 1) *Sports*: Data mining can be used by sports organizations in the form of statistical analysis, pattern discovery, as well as outcome prediction.
- 2) *Analyze Students Performance*: Information’s like Attendance, Class test, Seminar and Assignment marks were collected from the student’s management system, to predict the performance at the end of the semester.
- 3) *Telecommunication Industry*: Telecommunication services have integrated with the computer, internet, and network and with other communication technologies. Data Mining helps to identify telecommunications patterns, fraud activities and also helps to better use of resources and improve the quality of services.
- 4) *Retail Industry*: Data Mining plays a vital role in the retail industry also. Retail industry engrosses large amount of data that includes transportation, sales and consumptions of goods and services. E-commerce improves the online experience and describes the buying and selling of products, services, and information via computer networks including the Internet.

III. RESEARCH OBSERVATIONS

We have examined the publications, journals and reviews in the field of computer science and engineering, data mining and cardiovascular disease in recent times.

TABLE II.
SUMMARY OF MEDICAL DATA MINING TECHNIQUES.

REFERENCES	TECHNIQUES
[8],[11],[2],[4],[14], [3],[7],[13]	Neural Networks
[5]	Bagging
[15]	Decision Table
[5],[14],[7],[1],[6]	Naïve Bayesian
[2],[15],[14],[3],[7], [1],[6],[9],[5]	Decision Tree Algorithms such as ID3, C4.5, C5, and CART.

[2]	Bayesian Methods
[2]	SVM
[6]	REPTree
[6]	Bayes Net
[2]	Regression
[2],[4]	K-NN

IV. DATA MINING TECHNIQUES USED FOR PREDICTION

Different data mining classification techniques, i.e. Neural Networks, Decision Trees and Naive Bayes are used to analyse the dataset.

A. Decision Trees:

The decision tree approach is more powerful for classification problems. There are two steps in this techniques building a tree & applying the tree to the dataset. There are many popular decision tree algorithms CART, ID3, CHAID, and J48. From these J48, ID3, CART algorithms are used for diagnosis of heart disease.

- 1) J48 algorithm uses pruning method to build a tree. The J48 algorithm recursively classifies data until it has been categorized as perfectly as possible and gives maximum accuracy on training data. The overall concept is to build a tree that provides balance of flexibility & accuracy.
- 2) ID3 is the precursor to the C4.5 algorithm, and is typically used in the machine learning and natural language processing domains.
- 3) CART analysis is an umbrella term used to refer to both of the classification tree and regression tree and generates the binary decision tree.

B. Neural Networks:

An artificial neural network (ANN), often just called a "neural network" (NN), is a mathematical model or computational model based on biological neural network. In other words, it is an emulation of biological neural system. A Multi-layer Perceptron Neural Networks (MLPNN) is used.

C. Naive Bayes:

Naive Bayes classifier is based on Bayes theorem. This classifier algorithm uses conditional independence, means it assumes that an attribute value on a given class is independent of the values of other attributes. Naive Bayes is a statistical classifier and based on supervised learning.

V. CONCLUSION

The aim of this paper is to provide a study of different data mining techniques that can be helpful in heart disease prediction. Various techniques and data mining classifiers are defined in this work which has appeared in recent years for efficient and effective heart disease diagnosis. This paper has provided the summary of data mining techniques used for medical data mining.

From the analysis it is concluded that, data mining plays a major role in heart disease classification. The overall 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.

REFERENCES

- [1] **A.PRIYANGA and Dr.S.PRAKASAM** *The Role of Data Mining-Based Cancer Prediction system (DMBCPS) in Cancer Awareness* [Journal] // International Journal of Computer Science and Engineering Communications-IJCSEC.. - December 2013. - 1 : Vol. 1.
- [2] **Divya Tomar and Sonali Agarwal** *A survey on Data Mining approaches for Healthcare* [Journal]. - [s.l.] : International Journal of Bio-Science and Bio-Technology, 2013. - 5 : Vol. 5. - pp. 241-266.
- [3] **Beant Kaur and Williamjeet Singh** *Review on Heart Disease Prediction System using Data Mining Techniques* [Journal] // International Journal on Recent and Innovation Trends in Computing and Communication. - 10 : Vol. 2. - pp. 3003 – 3008. - ISSN: 2321-8169.
- [4] **Chandna Deepali** *Diagnosis of Heart Disease Using Data Mining Algorithm* [Journal] // (IJCSIT) International Journal of Computer Science and Information Technologies. - 2014. - 2 : Vol. 5. - pp. 1678-1680. - ISSN:0975-9646.
- [5] **Chaurasia Vikas** *Early Prediction of Heart Diseases Using Data Mining Techniques* [Journal]. - [s.l.] : Caribbean Journal of Science and Technology, 2013. - Vol. 1. - pp. 208-217.
- [6] **Hlaudi Daniel, Masethe Mosima and Anna Masethe** *Prediction of Heart Disease using Classification Algorithms* [Journal] // World Congress on Engineering and Computer Science(WCECS). - October 22-24, 2014. - Vol. II.
- [7] **K.Sudhakar and Dr. M. Manimekalai** *Study of Heart Disease Prediction using Data Mining* [Journal] // International Journal of Advanced Research in Computer Science and Software Engineering. - January 2014. - 1 : Vol. 4. - ISSN: 2277 128X.
- [8] **Miss. Chaitrali S. Dangare and Dr. Mrs. Sulabha S. Apte** *A Data Mining Approach for Prediction of Heart Disease using Neural Networks* [Journal]. - [s.l.] : International Journal of Computer Engineering & Technology (IJCET), October - December (2012). - 3 : Vol. 3. - pp. 30-40. - ISSN.
- [9] **Mohammad Taha Khan, Dr. Shamimul Qamar and Laurent F. Massin** *A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining* [Journal]. - [s.l.] : International Journal of Applied Engineering Research(IJAER), 2012. - Vol. 7. - ISSN 0973-4562.
- [10] **Mohammed Abdul Khaleel, Sateesh Kumar and Pradham G.N. Dash** *A Survey of Data Mining Techniques on Medical Data for Finding Locally Frequent Diseases* [Journal]. - [s.l.] : International Journal of Advanced Research in Computer Science and Software Engineering, August 2013. - 8 : Vol. 3. - ISSN: 2277 128X.
- [11] **Nidhi Bhatla and Kiran Jyoti** *An Analysis of Heart Disease Prediction using Different Data Mining Techniques* [Journal] // International Journal of Engineering Research & Technology (IJERT). - October - 2012. - 8 : Vol. 1. - ISSN: 2278-0181.
- [12] **Parvathi I and Siddharth Rautaray** *Survey on Data Mining Techniques for the Diagnosis of Diseases in Medical Domain* [Journal]. - [s.l.] : International Journal of Computer Science and Information Technologies(IJCSIT), 2014. - 1 : Vol. 5. - pp. 838-846. - ISSN:0975-9646.
- [13] **R. Chitra and V. Seenivasagam** *REVIEW OF HEART DISEASE PREDICTION SYSTEM USING DATA MINING AND HYBRID INTELLIGENT TECHNIQUES* [Journal]. - [s.l.] : ICTACT JOURNAL ON SOFT COMPUTING, July 2013. - 04 : Vol. 03. - ISSN: 2229-6956.
- [14] **Taneja Abhishek** *Heart Disease Prediction System Using Data Mining Techniques* [Journal] // ORIENTAL JOURNAL OF COMPUTER SCIENCE & TECHNOLOGY. - December 2013. - 6 : Vol. 4. - pp. 457-466. - ISSN: 0974-6471.
- [15] **Vikas Chaurasia and Saurabh Pal** *Data Mining Approach to Detect Heart Diseases* [Journal]. - [s.l.] : International Journal of Advanced Computer Science and Information Technology (IJACSIT), 2013. - 4 : Vol. 2. - pp. 56-66. - ISSN: 2296-1739.