

“PREDICTION OF HEART DISEASE USING DATA MINING TECHNIQUES” - A Review

Sonam Nikhar¹, A. M. Karandikar²

¹ M.Tech Student, CSE Department, Shri Ramdeobaba College of Engineering and Management, Nagpur, India

² Assistant Professor, CSE Department, Shri Ramdeobaba College of Engineering and Management, Nagpur, India

Abstract - In this paper, we present a critical review of the research now being undergoing in applications of data mining for a management of the healthcare system. The goal of this study is to explore emerging and new areas of data mining techniques used in healthcare management. One of the rapidly growing fields is health care industries. The medical industries have a great amount of data set collections about diagnosis, medications and patient details. To turns, these data is into a useful pattern and to predicting subsequent trends, data mining techniques are used in health care industries. The healthcare industries come across with new medication and medicine every day. The medical industries should offer better diagnosis and treatment of the patients to attaining a good quality of service. This paper explores different data mining techniques which are used in health care field for the prediction of heart diseases using data mining techniques.

Key Words: Data mining, Heart Disease Prediction, Classification algorithms.

1. INTRODUCTION

Data mining is the computer based process of extracting useful information from enormous sets of databases. Data mining is most useful in an exploratory analysis because of nontrivial information from large volumes of data.

The data mining tools are useful for answering business questions and techniques for predicting the various diseases in the healthcare field. Disease prediction plays an important role in data mining. There are multiple types of diseases predicting in data mining, namely heart diseases, breast cancer and lung cancer.

This paper analyzes the heart disease predictions using classification algorithms. These invisible patterns can be used for health diagnosis in healthcare data. Data mining technology affords an effective approach to the latest and indefinite patterns in the data. The information which is identified can be used by the healthcare

administrators to get better services. Heart disease was the most crucial reason for victims in the countries like India, United States. Data mining techniques like clustering, Association Rule Mining, Classification algorithms such as Decision tree [2], C4.5 algorithm, Neural Network [3], Naive Bayes [4] are used to explore the different kinds of heart - based problems [1]. Data mining methods like K-means clustering and C4.5 algorithm are used for validating the accuracy of medicinal data. These algorithms can be used to enhance the data storage for practical and legal purposes.

1.2. HEART DISEASE

The heart is an important organ of our body. If the operation of a heart is not proper, it will affect the other body parts of a human such as a brain, kidney etc. It is nothing more than a pump, which pumps blood through the body. If the circulation of blood in the body is inefficient the organs like brain suffers and if the heart stops working altogether, end occurs within minutes. Life is completely dependent on the efficient working of the heart. The term Heart disease refers to disease of heart & the blood vessel system within it.

There are number of factors which increase the chances of Heart disease:

- Hypertension
- Physical inactivity
- Obesity
- Poor diet
- High blood pressure
- High blood cholesterol
- Cholesterol
- Family history of heart disease
- Smoking

2. LITERATURE SURVEY

Numerous works in literature related to diagnosis of Heart disease using data mining techniques have

motivated this work. A brief literature survey is presented here.

A model Intelligent Heart Disease Prediction System (IHDP) built with the assistance of data mining techniques namely, Neural Network, Naïve Bayes and Decision Tree. Results show that each technique has its infrequent strength in realizing the objectives of the defined mining goals. IHDP is able to answer complex "what if" queries which conventional decision support systems cannot, was proposed by **Sellappan Palaniappan** et al. [2]. The results illustrated the uncouth strength of each of the methodologies in comprehending the goal of the specified mining objectives. IHDP was capable of responding queries that the traditional decision support systems were not able to. It facilitated the establishment of crucial knowledge such as patterns, relationships amid medical factors connected with heart disease. IHDP subsists well -being web-based, user-friendly, reliable, scalable and expandable.

The prediction of Heart disease, Blood Pressure and diabetes with the aid of neural networks was introduced by Niti Guru et al. [7]. Experiments were carried out on a sampled data set of patient's records. The Neural Network is trained and tested with 13 input variables such as Blood Pressure, Age, Angiography's report and the like. The supervised network has been advised for diagnosis of heart diseases. Training was carried out with the help of back propagation algorithm. Whenever unfamiliar data was inserted by the doctor, the system identified the unknown data from comparisons with the trained data and produce a catalog of probable diseases that the patient is vulnerable to.

In 2014, M.A.Nishara Banu B.Gomathy Professor, Department of Computer Science and Engineering has published a research paper "Disease Forecasting System Using Data Mining Methods"[8]. In this paper, the preprocessed data is clustered using clustering algorithms as K-means to cluster relevant data in a database. Maximal Frequent Item set Algorithm(MAFIA) is applied for mining maximal frequent patterns in heart disease database. The frequent patterns can be classified into different classes using the C4.5 algorithm as training algorithm using the concept of information entropy. The result demonstrates that the designed prediction system is capable of predicting the heart attack successfully.

In 2012, T.John Peter and K. Somasundaram Professor, Dept of CSE has presented a paper, "An Empirical Study on Prediction of Heart Disease using classification data mining technique"[5]. In this research paper, the use of pattern recognition and data mining techniques are used for prediction of risk in the clinical domain of cardiovascular medicine is proposed here. Some of the limitations of the traditional medical scoring systems are that there is a presence of intrinsic linear combinations of variables in the input set and hence they are not skilled at modeling

nonlinear complex interactions in medical domains. This limitation is handled in this research by use of classification models which can implicitly detect complex nonlinear relationships between independent and dependent variables as well as the ability to detect all possible interactions between predictor variables.

In 2013, Shamsheer Bahadur Patel, Pramod Kumar Yadav, and Dr. D. P. Shukla has presented a research paper, "Predict the Diagnosis of Heart Disease Patients Using Classification Mining Techniques" [6]. In this research paper, the health care industry, the data mining is mainly used for the prediction of heart disease. The objective of our works to predict the diagnosis of heart disease with a reduced number of attributes using Naïve Bayes, Decision Tree.

In 2012, Feixiang Huang, Shengyong Wang, and Chien-Chung Chan has presented a paper, "Predicting Disease By Using Data Mining Based on Healthcare Information System"[9]. In this paper, an author had proposed the data mining process to forecast hypertension from patient medical records with eight other diseases. Our focus on data mining is to extract hidden rules and relationships among diseases from a real world Healthcare Information System using Naive Bayesian and J-48 classifiers.

2.2 Data Source

A total of 603 records with 76 medical attributes (factors) all attributes are numeric-valued have obtained from the Cleveland Heart Disease database (cleveland.data). While the databases have 76 raw attributes, only 14 of them are actually used below we have lists the attributes.

1. **Age:** It will take age in years as input.
2. **Sex:** It will take two values as input i.e. value 1: Male and value 0: Female
3. **Chest Pain Type:** It will take four value as input which shows the chest pain type as value 1: typical type-1 angina, value 2: typical type angina, value 3: non-angina pain; value 4: asymptomatic.
4. **Trest Blood Pressure:** resting blood pressure (in mm Hg on admission to the hospital)
5. **chol:** serum cholesterol in mg/dl.
6. **Fasting Blood Sugar:** It will take two values as input i.e. value 1 for FBS > 120 mg/dl and value 0: for FBS < 120 mg/dl.
7. **Restecg:** Resting electrographic results will take 3 value as input i.e. value 0: normal; value 1: 1 having ST-T wave abnormality; value 2: showing probable or definite left ventricular hypertrophy.
8. **thalach:** Maximum heart rate achieved by the patient.
9. **Exang:** Exercise induced angina will take two values as input i.e. value 1: yes and value 0: no.
10. **Old peak:** ST depression induced by exercise relative to rest.

11. **Slope:** The slope of the peak exercise ST segment will take three values as input i.e. value 1: unsloping; value 2: flat and value 3: down sloping.
12. **CA :**Number of major vessels colored by fluoroscopy will take three value as input (value 0 – 3)
13. **Thal:** It will take 3 input values i.e. value 3: normal; value 6: fixed defect and value 7: reversible defect.
14. **num:** This a diagnosing attribute having two input value i.e. value 0: < 50% diameter narrowing (no heart disease); and value 1: > 50% diameter narrowing (has heart disease).

3. CONCLUSION

This paper presented a review of disease prediction for healthcare system using data mining techniques. Medical related information's are volumetric in nature and it could be derived from different birthplaces which are not entirely applicable in a feature. In this work, we have performed a literature survey on various papers. In future, we are planning to propose an effective disease prediction system to predict the heart disease with better accuracy using different data mining classification techniques such as Naïve Bayes and Decision Tree.

REFERENCES

- [1] V. Manikantan and S. Latha, "Predicting the analysis of heart disease symptoms using medicinal data mining methods", International Journal of Advanced Computer Theory and Engineering, vol. 2, pp.46-51, 2013.
- [2] Sellappan Palaniappan and Rafiah Awang, "Intelligent heart disease prediction system using data mining techniques", International Journal of Computer Science and Network Security, vol.8, no.8, pp. 343-350,2008.
- [3] K.Srinivas, Dr.G.Ragavendra and Dr. A. Govardhan," A Survey on prediction of heart morbidity using data mining techniques",International Journal of Data Mining & Knowledge Management Process (IJDKP) vol.1, no.3, pp.14-34, May 2011.
- [4] G.Subbalakshmi, K.Ramesh and N.Chinna Rao," Decision support in heart disease prediction system using Naïve Bayes", ISSN: 0976-5166, vol. 2, no. 2,pp.170-176, 2011.
- [5] T.John Peter , K. Somasundaram, "An Empirical Study on Prediction of Heart Disease using classification data mining technique" IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012) March 30, 31, 2012.
- [6] Shamsher Bahadur Patel, Pramod Kumar Yadav and Dr. D. P.Shukla, "Predict the Diagnosis of Heart Disease Patients Using Classification Mining Techniques",IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS),Volume 4, Issue 2 (Jul. - Aug. 2013).
- [7] Niti Guru, Anil Dahiya, Navin Rajpal, "Decision Support System for Heart Disease Diagnosis Using Neural Network", Delhi Business Review, Vol. 8, No. I (January - June 2007).
- [8] M.A.Nishara Banu, B.Gomathy, "Disease Forecasting System Using Data Mining Methods," International Conference on Intelligent Computing Applications,2014
- [9] Feixiang Huang, Shengyong Wang, and Chien-Chung Chan, "Predicting Disease By Using Data Mining Based on Healthcare Information System", IEEE International Conference on Granular Computing,2012.
- [10] Blake, C.L., Mertz, C.J.: "UCI MachineLearningDataset", <http://mllearn.ics.uci.edu/databases/heartdisease/>, 2004.