

Role of Different Data Mining Techniques for Predicting Heart Disease

Jyoti Thakur¹, Er. Munish Katoch²

¹Student, Dept of Computer Science Engineering, Sri Sai University Palampur (H.P) India

²Assitant Professor, Dept. of Computer Science and Engineering, Sri Sai University Palampur (H.P) India

Abstract – Heart is the important organ of human body. Changes in environmental conditions and lifestyle of people give rise to various diseases in humans related to heart and millions of people every year die because of heart problems. However several techniques have been suggested by various researchers in biomedical field to design the early prediction system for the heart related problems. The artificial neural networks (ANN), Genetic algorithm (GA), K-nearest neighbor (KNN) and many other classifiers of data mining were used to design the system for heart disease prediction. In this paper a detailed analysis and comparison of the various techniques used for the cardiovascular diseases (CVD) early prediction is done.

Keywords: Heart Disease Prediction, Data mining techniques, ANN, KNN, GA, CVD etc...

1. INTRODUCTION

The cardiovascular diseases are regarded as the high-ranking diseases and cause the millions of deaths every year all over the world. With the changing environment, and the growing old of population, these diseases will cause various difficulties for the human beings in following years [1]. Heart disease is the most common among all diseases which results in various disorders affecting the heart as well as blood vessels [2]. In traditional medical decision systems it was not possible to provide the intelligent decisions prior to the knowledge about the disease but now forecast of the diseases related to heart is possible and it is easy for the medical experts to give better and smart decisions. As per the report from World Health Organization (WHO), heart disease is the reason for 12 million deaths yearly, all over the world. 17.3 million Humans died in the year 2008, because of Heart problems. And WHO has made a prediction according to which around 23.6 million people may die because of heart problems [3]. Also the data mining schemes are used to extract the data from reports and results verify the affects of disease. This data mining plays a significant role in predicting the diseases in biomedical field. Nowadays a patient generally suffers from various diseases under the same category so it becomes quite uneasy for the doctor to diagnose the disease.

However the prediction about the diseases can be made by using the Data Mining having intelligent algorithms from the

data of patient having numerous inputs. Artificial neural network is usually used to deal with such difficult jobs. A lot of patient's data is used to train the neural network model; according to the past data of diseases of patients predictions about the diseases are made. The feed forward neural network is trained using back propagation. With passing time, this has establish itself as a standard scheme for the tasks related to classification and prediction about diseases in medical areas and others fields as well [4]. Many types of cardiovascular illness can be identified or diagnosed by taking into account family medical history and other variables. However, it was rather hard to predict heart disease without medical exams. The main of this paper was to diagnose various cardiovascular conditions and take any necessary precautions to avoid them at an affordable pace at an early point. In 'data mining' technology, characteristics for prediction of cardiovascular disease are fed into SVM and RF. This method was used in the preliminary measurements and surveys to detect cardiovascular disease at an early point and can be totally healed by appropriate diagnostics.[5]

In past some of the researchers tried to discover the optimal approach for risk prediction model. The survey and comparison is made on the basis of the researcher's discoveries and optimal model is selected. A literature survey of the several data mining techniques was done to foresee the heart disease is done in this paper. The survey make it is easy to know about the methods nowadays used to foresee the heart related diseases with the help of data mining classification of data.

2. DATA MINING TECHNIQUES

There is presence of various data mining schemes to know about the diseases related to heart. Some of the basic techniques from them are listed below [5].

2.1 Association:

It is the optimal data mining scheme among all the techniques. It is taken to predict the heart disease because it derive the relation between various features after analyzing; and helps the patient to know about the factors causing the heart disease and diagnosing the disease prior to its occurrence.

2.2 Classification:

Classification is the old data mining method and uses the machine learning. The classification is used to categorize the every item in the data set to previously defined groups. This classification model use the mathematical operations like decision trees, neural network and linear programming

2.3 Clustering:

It takes the help of automatic method to make the cluster from the data having same types of features. These clusters provide the significant information about the data. Clustering characterize the classes and insert objects to them. On the other hand there are previously defined classes in case of classification and objects are allocated to those classes. For example, predicting the disease using clustering. When all patients have similar disease and equal risk factor then clustering arrange the data depending on the various parameters like high blood pressure, sugar level etc.

2.4 Prediction:

This is one of the data mining scheme which determine the relation among independent variables and dependent variables correspondingly.

In terms of market, prediction about the profit can be made in advance for future reference if we assume that sale is independent, while the profit can be made dependent also. So as per the past sale data and profit regression graph can be drawn to make assumption for the future profit. The various data mining techniques are summed up as below

Table -1: Techniques of Data Mining

S. No.	Data Mining Task	Data Mining Algorithms & Techniques
1.	Classification	Decision Trees, Rule-based, Neural Networks, Naïve Bayes and Bayesian Belief Networks, Support Vector Machines, Genetic Algorithms
2.	Clustering	K-Means
3.	Regression and Prediction	Support Vector Machines, Decision Trees, Rule induction, NN
4.	Association and Link Analysis (finding correlation between items in a dataset)	Association Rules Mining (ARM)
5.	Summarization	Multivariate Visualization

3. OTHER RELATED WORK

A lot of research is done by the researchers to make the heart related systems better like prediction about disease and several data mining techniques have been used for this purpose. All the efforts are done to make the system more accurate and efficient to know about the occurrence of heart attack in advance. Several data mining methods used by researchers are analyzed in this paper. The data mining schemes are used for detecting the cardiovascular diseases for various datasets related to heart disease.

The study [1] about the five classifiers models is done here which are used to detect and guess about occurrence of CVD. These five classifier models are: Regression analysis, Naïve Bayesian, Artificial neural network, random forest and vector machine. The CVD data was extracted from the from the Cleveland and Hungary UCI datasets. The results from the experiments verify that the hybrid model was a better in offering the high accuracy about the prediction and reliable. From the study a better, smart, economical, GUI friendly interface system can be design in the future.

The main objective of using data mining is feasibility about the health [2] disease. The main motive of our research work is to early diagnosis of the heart disease by reducing the parameters used for this process. Fourteen parameters were used to diagnose the heart ailment. Now the number is decreased to 6 from 14 with the help of Genetic algorithm. After decreasing the parameters previously explained classifiers i.e. Naïve Bayes, Clustering and Decision tree are used to envisage the heart ailment.

[7] Author designed an upgraded system to predict the heart ailment with the help of classification schemes using data mining. Author deployed increased number of features as inputs and analyzed prediction system. The model used by author took attributes like blood, sex, blood pressure to foresee the chances for heart attack. This also includes the two characteristics i.e. obesity and smoking. The Naïve Bayes, artificial neural network and decision tree classifiers were used to make the analysis of heart related diseases. A comparison of various parameters for different techniques with the designed technique was done. The system was 100% precise in case of neural network, 90.74% for the Naïve Bayes and for the decision tree the precision was 99.62%.

Author of [8] designed a system for the heart related problem verdict using classification model. The open access data was divided by k-means technique and here K= 2 for testing rule of cluster analysis. The clusters were made using dataset related to heart diseases and the classification rule scheme in data mining used to analyze the system. The classification rule scheme under data mining which is also

termed as projective adaptive resonance theory is analyzed on the basis of the related dataset. For the assessment of unbiased total heart prediction system 10 folding cross validation technique was used. The precision values for the K-means clustering PART based k means and PART clustering scheme are 81.08%, 84.12% and 79.05%. From the results it was found that classification-dependent clustering was better in detecting heart related diseases and was more precise.

In year 2004, Y. Alp Aslandogan, et. al. [9] performed a work, "Evidence Combination in Medical Data Mining". The different aspects of three classifiers i.e. K-Nearest Neighbor (K-NN), Decision Tree and Naïve Bayesian are shown in paper. Different features and aspects of three classifiers were merged using Dempster's rule and on the basis of that a conclusion was made. An experiment using k-fold cross validation done. The experiment shows that the behavior of data sets used was improved for some of the classifiers while not for others. The results of the combined classifiers idea confirmed that the accuracy was better than the other systems. The performance of Dempster combination approach and performance based linearity and majority selection combination model shows that the combination provides better classification.

. In Year 2010, Harsh Vazirani, et. al. [10] performed a work, "Use of Modular Neural Network for Heart Disease". This paper mainly focused on the heart ailment early detection. The two diagnosis methods for heart disease were used. One method is manual while the second one automatically determines the disease by taking the help of intelligent and smart expert system. Modular neural network was also used for diagnosing heart related issues. The feature were classified and assigned to two different neural network model which are Back propagation (BPNN) and Radial Basis Function Neural Network (RBFNN). These models were used for the purpose of training and testing.

In, [11] The genetic algorithm and KNN approaches were taken into account to address the issue of heart ailment and improve the accuracy. The genetic algorithm here mainly was used to reduce the undesirable parameters and choose only those which are helpful in classification process. Author trained the KNN classifier to make the groups of heart ailment depending on the dataset either of healthy or of sick person. The performance of this technique was measured with 6 medical datasets and 1 dataset which were not from medical field. And the result showed that the combination of genetic algorithm and KNN significantly enhance the output of system

Author in [12], used K-nearest neighbor classifier and fuzzy system to make the prediction related to any heart disease. The KNN algorithm was used to improve the accuracy for the

detecting heart disease. The KNN is better when used with the fuzzy algorithm for some of the diseases. The KNN also provide more quality for datasets.

Author in this paper have focused on the detection of the heart related ailment, blood pressure, and some other malady using neural networks. Niti Guru et.al [13] performed experiments on the raw data set of sick patient records. Using patient's data the neural network model was trained. Different samples like Blood pressure, age etc. variables were tested with the model. The network model was used for identifying the heart diseases. The back propagation model was used to train the model. When the data which was not related to model, the system detected the irrelevant data and comparison with the trained model and list of diseases which could be caused because of any particular parameter was made.

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"[14]. The pattern recognition and data mining schemes were used to make prediction of danger of CVD in this paper. There was some limitations old medical scoring mechanism. The limitations include the presence of intrinsic linear combination variables as the inputs and they were not that much master in designing the model of nonlinear complex works in medical area. The classification model was used to manage these problems. This model can diagnose the complex relations between dependent and independent variables.

Chaitrali S. Dangare and Sulabha S. Apte [15] in this paper proved that the artificial neural network works better than the other data mining schemes like Naïve Bayes and Decision Tree. The heart disease forecast system was designed with the help of 15 parameters as input to system. The two additional parameters smoking and obesity were inserted to make the system of heart prediction system efficient.

As the artificial neural network k (ANN) has high accuracy and learning speed, so it can be used in the medical field of heart related problem detection [16]. Kumaravel et.al have suggested an automatic diagnosis system ailments related to heart which uses neural network. The features from the ECG of patient were taken out. 5 major heart diseases were classified using various 38 parameters. The accuracy around 63.6 to 82.9% [17] was achieved using the suggested method.

B. Venkatalakshmi and M.V Shivsankar in year 2014 [18] done a research on the heart diseases prediction with the help of Naïve Bayes and Decision Tree data mining methods. Number of times the same experiment was performed for the same data. Software WEKA 3.6.0 was used for this

simulation purpose. Data set from 294 records of patients used to perform the experiment. 13 different parameters were used and it was found that Naïve Bayes technique works better than the Decision tree scheme and had higher efficacy.

In 2013, Shamsher 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" [19]. Data mining technique was mainly used for heart related problem detection. The main focus of researchers here was to reduce the number of parameters used for detection of heart disease in advance. The naïve bayes and decision tree classifiers were used for this purpose.

4. CONCLUSION

According to a report, millions of people die every year just because of heart diseases. This situation is becoming worst with the changes in the environment and lifestyle of people. However if the prediction about the disease may be made earlier then lot of lives can be saved. For this data mining is playing an important role in early prediction of diseases. Data mining is used to extract the useful information from the datasets of patient which helps in determining the risk factor for heart disease and designing the system for heart disease prediction. The different classifiers used for classification of the dataset are artificial neural network (ANN), K nearest neighbor (KNN), Genetic algorithm (GA), Naïve bayes and decision tree whose literature survey is done in this paper. The comparison in terms of accuracy, efficiency and parameters used in feature extraction was done and it was found that ANN uses less parameter and provides better accuracy for heart prediction system with the same dataset used for other classifiers. As a future scope ANN can be modified or can be improved to achieve effective prediction. Also classifiers from same class can also be helpful and can provide much improved prediction systems to save number of people.

REFERENCES

- [1] Jan, M., Awan, A. A., Khalid, M. S., & Nisar, S. (2018). Ensemble approach for developing a smart heart disease prediction system using classification algorithms. *Research Reports in Clinical Cardiology*, Volume 9, 33-45. doi:10.2147/rcc.s172035
- [2] Shamsher Bahadur P. , Pramod Kumar Y., et al., "Predict the Diagnosis of Heart Disease Patients Using Classification Mining Techniques", *IOSR Journal of Agriculture and Veterinary Science* , Volume 4, Issue 2 (Jul. - Aug. 2013), PP 61-64
- [3] Miss. Chaitrali S. Dangare, et. al, "A data mining approach for prediction of heart disease using neural networks", *international journal of computer engineering*
- [4] R. Chitra and V. Seenivasagam, "Review Of Heart Disease Prediction System Using Data Mining And Hybrid Intelligent Techniques" *ictact journal on soft computing*, july 2013, volume: 03, issue: 04
- [5] Mamatha Alex P and Shaicy P Shaji, "Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique".
- [6] V. Krishnaiah, G. Narsimha et. al, "Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review" *International Journal of Computer Applications (0975 - 8887) Volume 136 - No.2, February 2016*
- [7] S. S. Dangare, C. S and Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques," *Int. J. Comput. Appl.*, vol. 10, no. 10, pp. 44-48, 2012.
- [8] A. Kumar, P. Prabhat, and K. L. Jaiswal, "Classification Model for the Heart Disease," *Glob. J. Med. Resarch Dis.*, vol. 14, no. 1, 2014.
- [9] Y. Alp Aslandogan et. al., "Evidence Combination in Medical Data Mining", *Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04) 0-7695-2108-8/04©2004 IEEE*.
- [10] Harsh Vazirani et. al., "Use of Modular Neural Network for Heart Disease", *Special Issue of IJCCT Vol.1 Issue 2, 3, 4; 2010 for International Conference [ACCTA-2010], 3-5 August 2010, page no. 88-93.*
- [11] M. Akhil, B. L. Deekshatulu, and P. Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm," *Procedia Technol.*, vol. 10, pp. 85-94, 2013.
- [12] A. A. Chaudhari, "Fuzzy & Datamining based Disease Prediction Using K-NN Algorithm," *Int. J. Innov. Engineering Technol.*, vol. 3, no. 4, pp. 9- 14, 2014.
- [13] A. A. Chaudhari, "Fuzzy & Datamining based Disease Prediction Using K-NN Algorithm," *Int. J. Innov. Engineering Technol.*, vol. 3, no. 4, pp. 9- 14, 2014
- [14] Niti Guru, Anil Dahiya, Navin Rajpal, "Decision Support System for Heart Disease Diagnosis Using Neural Network", *Delhi Business Review*, Vol. 8, No. 1 (January - June 2007).
- [15] 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*
- [16] Chaitrali S. Dangare, Sulabha S. Apte, -Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques; *International Journal of Computer Applications (0975 - 888) Volume 47- No.10, June 2014.*

- [17] Nidhi Bhatlet and Kiran Jyoti, "An Analysis of Heart disease prediction system using different data mining techniques", International Journal of Engineering Research and Technology, ISSN, volume 1, October-2016
- [18] N. Kumaravel, K. Sridhar, and N. Nithiyandam, "Automatic diagnoses of heart diseases using neural network", In Proceedings of the Fifteenth Biomedical Engineering Conference, pp. 319-322, March 2016
- [19] B.Venkatalakshmi, M.V Shivsankar, Heart Disease Diagnosis Using Predictive Data mining; International Journal of Innovative Research in Science, Engineering and Technology Volume 3, Special Issue 3, March 2016
- [20] Shamsheer 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 (IOSRJAVS), Volume 4, Issue 2 (Jul. - Aug. 2013).