A Review on Intutive Prediction of Heart Disease Using Data Mining Techniques

Akansha Jain, Manish Ahirwar, Rajeev Pandey

IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-7, Issue-7, July 2019 E-ISSN: 2347-2693

A Review on Intutive Prediction of Heart Disease Using Data Mining Techniques

Akansha Jain[1*], Manish Ahirwar[2], Rajeev Pandey[3]

[1,2,3]Department of Computer Science, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, India

[*]*Corresponding Author: jainakansha1000@gmail.com,, Tel.: +91-7879780944*

DOI: https://doi.org/10.26438/ijcse/v7i7.109113 | Available online at: www.ijcseonline.org

Accepted: 16/Jul/2019, Published: 31/Jul/2019

Abstract— Healthcare evaluates clinical datasets regularly by specialist's learning and action. In the clinical field, computer-supported with prediction system is used in the healthcare department. Data mining approach provides innovation and strategy to replace voluminous information into useful data for achieving a decision. By utilizing information mining systems it needs less investment for the forecast of the sickness with more accuracy and precision. This paper evaluates various classifiers and algorithms are used for the expectation of cardiovascular illness.

Keywords— WEKA tool, Data Mining techniques, Heart disease prediction, Computer Aided Support System.

I. INTRODUCTION

Most fundamental hard-working muscular organ of our body is heart. In circulatory system, blood is transferred through the veins in heart. This muscular system plays a vital role as it transports oxygen, blood and other materials to the various body parts [1]. It might cause serious wellbeing conditions including death if the heart does not work appropriately. It results in several illness, disability and death. Modifiable hazard factors incorporate weight, smoking, absence of physical movement, etc. Illness diagnosis plays a leading role in clinical field. Intelligent data mining algorithms tackle problem of clinical dataset prediction involving several inputs.

Decision support systems with computer-based information and can assist in accomplishing health related tests at a decreased expense. Computerized system needs a relative study of different techniques available for exact and efficient execution. This paper predicts numerous coronary illness prediction by utilizing the approaches of data mining proposed in recent years.

Section I contains the introduction of data mining algorithms, data mining tools and cardiovascular diseases, Section II contains the related work of prediction system using different datasets using different approaches, Section III contains the conclusion.

A. DATA MINING ALGORITHMS

Various algorithms formulates due to different research works on data mining. These techniques are straightforwardly utilized for developing frameworks or to find crucial inferences and conclusions from the resulted dataset. Various well-known techniques are Support vector machine, K-means, Naïve Bayes, Artificial neural network etc are discussed.

a. Decision Support

This tool utilizes a tree like model or graph of decisions so that possible results which include chance event outcomes and utility. This tool is one of the common approach to show algorithms [2]. These are generally utilized for research operations more specific in analysis of decisions to help and evaluate a procedure that can achieve the objective. It is most useful and common tool in machine learning. One by one mapping from root nodes to leaf node so that this can effortlessly be changed in the form of rules and at last by following these norms, appropriate results can be found.
This tool utilizes a tree like model or graph of decisions so that possible results which include chance event outcomes and utility.

b. Naïve Bayes Algorithm

A simplest technique for building classifiers and a probabilistic classifier depends on Bayes' theorem. Bayes theorem is calculated as follows: $P(C|X) = P(X|C) * P(C) / P(X)$, where C is the class such that $P(X)$ is constant for all classes and X is data tuple. All classifiers suppose that the value of any particular feature is independent of the value of any other feature, given the class variable [3]. In spite of the fact that it suppose an unrealistic condition that attribute values are conditionally independent, it performs surprisingly well on voluminous datasets where this condition is assumed and holds.

c. Support Vector Machine

SVM is used for accessing information and pattern for regression analysis and classification. The objective is to find most accurate classification function to recognize individuals from the two classes in the training dataset. It depends on mathematical functions and utilized in model complex, and real world problems. It works well on data sets that have many attributes and ensures that the best such function is found by maximizing the margin between the two classes. It differentiates data by searching the best hyper plane that divides all data points of one class from those of the other class.

B. DATA MINING TOOLS

Tools provide ready applications to be used for mining algorithms. They have an easy to use interface and researchers can easily use them because of free open source software. Popular data mining tool is WEKA, ORANGE, MATLAB etc.

a. WEKA Tool

WEKA tool is created in New Zealand by University of Waikato which uses Java language consists of various data mining algorithms. This tool perform data mining tasks due to the collection of machine learning techniques which are applied directly on datasets. Data preprocessing, classification, clustering, regression, visualization tools, association rules are provided by WEKA tool [4]. This tool is an open source software where ARFF file format is used by WEKA which identify different things using special tags.

C. CARDIOVASCULAR DISEASE

Cardiovascular ailment is one of the primary reasons of death for both men and women. The term coronary illness identifies with various ailments identified with heart .So irregular health conditions are defined that directly stimulate the heart and all its parts.

Various heart related cardiovascular diseases along with description are given in Table 1.

Table 1 - Types of Cardiovascular Diseases

Heart-related cardiovascular diseases	Description
Heart failure	Heart muscle doesn't pump blood
Angina	Chest pain due to a lack of blood to the heart muscle
Arrhythmia	Atypical heart rhythm
Cardiomyopathy	Heart muscle disease
Congenital heart disease	Heart disfigurements that are present at birth
Coronary heart disease	Arteries supplying blood to heart muscle becomes

II. RELATED WORK

- R. Ansari et al [1] reviewed data mining on heart disease prediction and diagnosis. This study tries to support healthcare specialists to early analyze cardiovascular disease and evaluate related hazard factors. At last the main cardiovascular disease diagnosis indices were recognized using expert's belief. Then, data mining techniques were applied on a cardiovascular related dataset. Finally, cardiovascular illness attributes were identified and a model was developed based on extracted rules and classifiers. Visual Studio had been used to write the coding of algorithm.

- Mirpouya Mirmozaffari, Alireza Alinezhad and Azade Gilanpour et al [5] reviewed data mining classification algorithm for cardiovascular disease prediction for getting better results, clinical datasets are preprocessed using various unsupervised and supervised algorithms. Various methods of evaluation are utilized to find best algorithm and different prediction model contains different filters, analysis models. WEKA tool is used where 396 approaches and 9 stages are compared using Multilayered filtering preprocess for accuracy. Random tree technique preforms best and results an accuracy of 97.6077% with least errors.

- G. Purusothaman and P. Krishnakumari et al [4] proposed an audit on information mining techniques for prediction of cardiovascular disease. Various methods are applied to evaluate heart disease risk prediction. By using classification techniques, heart disease prediction model find results at least expense and exertion. Following are the two prediction models used for data analysis. Applying single model to voluminous model related to heart data is first one and second is for implementing combined model to get hybrid model.

- Kiruthika Devi, D. Kalita et al [6] proposed in health care domain to survey numerous data mining tools for heart disease prediction and diagnosis. A coronary illness expectation model uses algorithms of data mining to support doctors in predicting cardiovascular disease result on the basis of clinical information. In the field of medicinal services, data mining classification techniques like SVM, Naïve Bayes,

decision tree etc. helps in better decision making. Algorithms gathering assists to make decisions more accurate and quicker.

- Vincy Cherian et al [3] proposed prediction using Laplace smoothing and Naïve Bayesian algorithms. The prediction model is assumed to keep away irrelevant diagnosis test conducted and delay in starting right treatment .Therefore, by early diagnosing patient with heart disease both time and money can be saved. Therefore, doctors can easily diagnose without unnecessary treatment provoked because of inexperience and intuitions made by doctor. This paper additionally looks at the exactness when attributes are decreased for prediction. A voluminous training dataset would definitely provide appropriate results which can train the system model properly. The framework can be extended to incorporate different fields for various ailment expectation.

- S. Immamul Ansarullah, Pradeep K. Sharma et al [7] evaluated a study on cardiovascular illness prediction using various algorithms of data mining to overcome the problem by analyzing hidden patterns which results in useful information from complex and voluminous data. It highlights the vital role played by data mining tools in analyzing huge volumes of healthcare related data in prediction and diagnosis of disease. The different Data mining tools can help an expert in effective decision making and better diagnosis in healthcare sector. Neural Networks based techniques with 15 attributes has performed best in comparison with Decision Tree based techniques which showed better performance of 99.62%.

- Sushmita Manikandan et al [8] proposed heart attack prediction system using machine learning repository of UCI to develop model. Rapid Miner is utilized for purifying the dataset and Anaconda v2.7 bundles to construct the classifier. Naïve Bayesians theorem was used to develop the classifier and GUI is web based. A model of framework which classifies risk factor on an individual is developed. Process of analysis and prediction can be successfully completed using different tools of data mining. A prototype developed evaluates relationship between the buried patterns and makes a prediction.

Table 2: Comparison between Heart Disease Prediction System using Data Mining Techniques

Attributes	**Mirpouya Mirmozaffari** [5]			**H. Benjamin Fredrick** [9]			**Shan Xu** [10]		
Technique Used	**Random Tree**	**IBK**	**Random Forest**	**Naïve Bayes**	**Decision Tree**	**Random forest**	**SVM**	**Random Forest**	**Bayes**
TP Rate	97.6%	97.6%	97.6%	57.1%	69.8%	74.6%	98.9%	1.1%	54.1%
FP Rate	2.1%	2.1%	2.1%	57.1%	32.8%	28.3%	1.1%	3%	45.4%
Precision	97.7%	97.7%	97.7%	32.7%	69.6%	74.5%	98.9%	97.1%	54.9%
F-measure	97.6%	97.6%	97.6%	41.6%	69.6%	74.3%	98.9%	97.0%	53.0%
Recall	97.6%	97.6%	97.6%	57.1%	69.8%	74.6%	98.9%	97.0%	54.1%
Objective	Discovery of unfamiliar patterns in cardiovascular diseases, easily available algorithms of classification are referred to dataset and their accuracies are compared.			To find the best classification algorithm applied for providing best accuracy whether person is suffered with heart disease or not.			To establish practical prediction system based on data mining techniques with more accurate results to medical service.		
Environment Used	WEKA			NA			NA		
No. of attributes	8			13			13		
Data set	Clinical dataset			Statlog dataset			Cleveland Database		
Result	Accuracy of 97.6077% with random tree and considered as lowest error with highest performance.			Random Forest algorithm with 81% precision results best as comparison to other algorithms for heart disease prediction.			Results 97% accuracy which is better as compare with other classifiers except SVM(98.9%).		

Table 3: Comparison between Heart Disease Prediction System using Data Mining Techniques

Attributes	**Ramin Assari** [1]			**Vincy Cherian** [3]	**Sushmita Manikandan** [8]
Technique Used	**Decision Tree**	**Naïve Bayes**	**SVM**	**Naïve Bayes classification**	**Gaussian Naïve Bayes algorithm**

Accuracy	79%	83.66%	84.44%	86%	81.25%
Specificity	81.48%	86.41%	89.5%	96.1%	83.67%
Sensitivity	76.08%	80.43%	78.26%	75%	NA
Objective	An accurate and systematic mode for diagnosis and prediction of heart disease victim.			To develop tool for distinguishing patients with heart disease which provides training tool to nurses and medical students.	The main aim is to reduce time and efforts done for prediction by binary classifier.
Environment Used	Weka and IBM SPSS Modeler			NA	Python, web interface
No. of attributes	13			[11] [11]13	14
Data set	Heart disease dataset of Cleveland Foundation.			Cleveland's heart disease datasets.	UCI's machine learning repository
Result	Achieved highest accuracy (84.33%) with SVM technique.			Proposed model is more accurate since it has 86% accuracy with 13 attributes.	Naïve Bayes gives the best result with accuracy of 81.25%.

III. CONCLUSION

Over the world huge reason for death is heart diseases. On the other hand, their early diagnosis helps in improving patients' health status and results in life-saver. Therefore, this study of papers focused to assist physicians in early diagnose and analyze heart disease risk factors using data mining techniques. This paper investigates different experiments from various surveys that are conducted to find the most accurate classifiers so that heart diseases can be predicted as soon as possible.

The tools used in data mining assists in the process of diagnosis and prediction which in turn assess lower complications and costs on coronary patients. After evaluating various surveys and research papers, picking diverse data mining strategies and actualizing them on the chose dataset, highest accuracy results in SVM technique. The performance of Naive Bayes shows high level compare with other classifiers except SVM technique.

REFERENCES

[1] R. Ansari, P. Azimi and M. R. Taghva, "Heart Disease Diagnosis Using Data Mining Techniques," 2017.

[2] M. A. Karaolis, "Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining with Decision Trees," *IEEE Transcations on Information Tehnology in Biomedicine.*, 2010.

[3] V. Cherian and B. M.S, "Heart Disease Prediction using Naive Bayes Algorithm & Laplace Smoothing technique," *International Journal of Computer Science Trends & Technology,* vol. II, no. 2, pp. 68-73, 2017.

[4] G. Purusothaman and P. Krishnakumari, "A Survey of Data Mining Techniques on Risk Prediction:Heart Prediction," *Indian Journal of Science and Technology,* vol. 8, no. 12, pp. 2-5, 2015.

[5] M. Mirmozaffari, A. Alinezhad and A. Gilanpour, "Data Mining Classification Algorithm for Heart Disease Prediction," *International Journal of Computing,Communication & Instrumentation,* vol. IV, no. 1, 2017.

[6] S. Kiruthika Devi, S. Krishnapriya and D. Kalita, "Prediction of Heart Disease using Data Mining Techniques," *Indian Journal of Science and Technology,* vol. 9, no. 39, pp. 1-5, 2016.

[7] S. I. Ansarullah, P. k. Sharma, A. Wahid and M. M. Kirmani, "Heart Disease Prediction System using Data Mining Techniques: A study," *International Research Journal of Engineering and Technology,* vol. III, no. 08, pp. 1375-1381, 2016.

[8] S. Manikandan, "Heart Attack Prediction System," *International Conference on Energy,Communication,Data Analtics & Soft computing,* pp. 817-820, 2017.

[9] H. B. F. David and S. A. Belcy, "Heart Disease Prediction Using Data Mining Technique," *ICTACT,* vol. IX, no. 1, pp. 1817-1823, 2018.

[10] S. Xu, Z. Zhang, D. Wang and J. Hu, "Cardiovascular Risk Prediction Method Based on CFS subset Evaluation & Random Forest Classification," *IEEE 2nd International Conference on Big Data,* pp. 228-232, 2017.

[11] M. N. Kumar, K. V. S. Koushik and K. Deepak, "Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology,* vol. III, no. 3, pp. 888-898, 2018.

[12] N. I. Patel and H. R. Patel, "A Survey on Prediction of Disease with Data Mining," *International Journal of Computer Sciences and Engineering,* vol. VII, no. 2, pp. 289-

293, 2019.

[13] P. Singh, S. Singh and G. S. P. Jain, "Effective heart disease prediction system using data mining techniques," *International Journal of Nanomedicine*, 2018.

[14] V. Chaurasia and S. Pal, "Data Mining Approach to Detect Heart Disease," *International Journal of Advanced Computer Science and Information Technology*, 2014.

[15] J. Vijayashree and N. C. SrimanNarayanaIyengar, "Heart Disease Prediction System Using Data Mining and Hybrid Intelligent Techniques: A Review," *International Journal of Bio-Science and Bio-Technology*, vol. VIII, no. IV, pp. 139-148, 2016.

[16] V. Krishnaiah, G. Narsimha and N. S. Chandra, "Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review," *International Journal of Computer Applications*, vol. 136, no. 2, pp. 43-51, 2016.

[17] R. Thakur , B. Gauri, D. Ashwini and . P. Shinde, "Heart Attack Prediction System Using Data Mining," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. IV, no. 8, pp. 15582-15585, 2016.

[18] A. Hazra, S. K. Mandal, A. Gupta, A. Mukherjee and A. Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques:A Review," *Advances in Computational Sciences and Technology*, vol. X, no. 7, pp. 2137-2160, 2017.

[19] P. S. C. Reddy, P. Palagi and Jaya, "HEART DISEASE PREDICTION USING ANN ALGORITHM IN DATA MINING," *International Journal of Computer Science and Mobile Computing*, vol. VI, no. 4, pp. 168-172, 2017.

[20] K. T K, "Heart Disease Prediction System using Data Mining Method," *International Journal of Engineering Trends and Technology*, vol. 47, no. 6, pp. 361-363, 2017.

[21] E. S. Kajal, "Prediction of Heart Disease using Data Mining Techniques," *International Journal of Advance Research , Ideas and Innovations in Technology*, vol. II, no. 3, pp. 1-7, 2016.

[22] A. Kaur and J. Arora, "HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES: A SURVEY," *International Journal of Advanced Research in Computer Science*, vol. XI, no. 2, pp. 569-572, 2018.

[23] M. Raihan, S. Mondal, a. More, P. K. Boni and M. O. F. Sagor, "Smartphone Based Heart Attack Risk Prediction System with Statistical Analysis and Data Mining Approaches," *Advances in Science, Technology and Engineering Systems Journal*, vol. II, no. 3, pp. 1815-1822, 2017.

[24] K. Gomathi and D. Shanmugapriyaa, "Heart Disease Prediction Using Data Mining Classification," *International Journal for Research in Applied Science & Engineering Technology*, vol. IV, no. II, pp. 60-63, 2016.

[25] A. Hazra, S. K. Mandal, A. Gupta, A. Mukherjee and A. Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques:," *Advances in Computational Sciences and Technology*, vol. X, no. 7, pp. 2138-2158, 2017.

Authors Profile

Akansha Jain is currently pursuing Dual Degree Integrated Post Graduation in Department of Computer Science and Engineering, University Institute of Technology RGPV, Bhopal, (M.P.) since July 2014. Her research area is based on prediction of heart disease using data mining techniques for the welfare of healthcare department.Implementation and design of clinical decision support to assist doctors and professionals in diagnosis and prediction. This model goes through evaluation of knowledge base and clinical database of cardiovascular patients.

Prof. Manish Ahirwar is an Assistant Professor in Department of Computer Science and Engineering, University Institute of Technology RGPV, Bhopal, (M.P.) since july 2007. He has 12 years of academic experience. He received his Bachelor's degree in Computer Science and Engineering in the stream of Information Technology. He has done Ph.D from University Institute of Technology RGPV, Bhopal, (M.P.) in stream of computer science. He is famous for academic, administrative and motivational skills. His motive is to spread practical knowledge to develop students and institute as a whole.

Dr. Rajeev Pandey is an Assistant Professor in Department of Computer Science and Engineering ,University Institute of Technology RGPV, Bhopal (M.P.) since july 2007. He has 12 years of academic experience. He received his Bachelor's degree in Computer Science and Engineering from IET ,DR. B.R.A. University, Agra (U.P.). He has done M.E. in Computer Science and Engineering in 2004 & Ph.D in 2010 from DR. B.R.A. University, Agra (U.P.), India

