

Heart Disease Prediction using Machine Learning Techniques Comparative Study

Sharmila Zope¹, Nandini Chaudhari², Nita Patil³

School of Computer Science KBCs North Maharashtra University, Jalgaon, India^{1,3}

DRS Kiran and Pallavi Patel Global University, Vadodara, India²

Corresponding author: Sharmila Zope, Email: sharmilazope@gmail.com

Now a days there is a significant amount of information is present in across the internet in the form of various types more specifically the information on, websites, news, blogs and other digital content. But there is valuable or meaningful information which is hidden inside the data which is very crucial for taking many important decisions. Thus this research is very useful for obtaining such useful information from the available one. The best tool for the extracting the useful information from the available large amount of information is known as data mining. This research deals to get the useful information from large amount of data and which is used in taking the crucial decision. Data mining is the one of the important tool to extract useful and meaningful information from the available large amount of data. Hence data mining is used in most of the applications like healthcare, whether forecasting and entertainment. The importance of data mining in the field of healthcare has proven its importance particularly in preventing, predicting and detecting and also in curing most of the heart diseases should be considered as milestone.

Keywords: Classification Techniques, Electrocardiogram Analysis, Machine Learning, Heart Disease.

1 Introduction

Data mining deals with extraction and discovery of useful data pattern from the available huge amount of information. This process of data mining is accomplished with many new techniques such as artificial intelligence, machine learning and most of the specialized algorithms which are very much application specific. Such as healthcare military and weather forecasting and it industry where a large amount of information has to be processed on the daily basis and drawing the useful conclusion from the such a large data is very essential in order to take the proper decisions. Now a day's most of the private companies also make use of the data mining techniques for getting a huge hike in the profit. By applying such techniques in the medical sector particularly for detection of heart diseases has opened a new dimension in the history of the medical and computer science where the life of the people can be saved using data mining techniques. This research is very useful in the focusing on different dimensions of analysis, diagnosis or prediction of the disease by using different tools and techniques.

There are many reasons for the heart attack including life style and lack of exercise or may be the wrong eating habits. This may include the other reasons also. Doctors, medical practitioners and health care professionals are also continuously working over the finding the new causes of the heart diseases on the other hand they are also working on finding the new methods of detecting and curing of heart diseases.

As the population of heart patients are increasing day by day in the whole world it is has become a prime concern to resolve this issue. Lots of research and studies are continuously being done all over the world for early detection and best treatment after detection of heart diseases. This research focuses on various aspects of not only reducing the risk of heart diseases but also curing them at the best. And this all can be done with the help of data available however the data for this kind of research is available in huge amount and is very difficult and is very tedious to take decision based on such a data. Here the role of data mining comes in to play which help in extracting very useful information from the available large amount of data. [2].

Medical decision support systems are specially designed for proper and perfect diagnosis of the diseases and their correct treatment. Such systems are very important in taking the crucial decisions regarding the study, research and treatment of different diseases which consist of the huge amount of data in different forms like text, images like x-ray, ECG and MRI, videos or sound or it may be in the form of any sort of multimedia data and that too in huge amount. The improvements in the medical decision can be significantly improved by increasing the quality of such support systems which deals with analysis of medical data and different types of knowledge of this particular domain Data mining helps in extracting the useful knowledge. These data patterns may include traditional or statistical approaches for analyzing the kind of data available. Ongoing collaborations between cardiovascular clinicians and computer scientists are watching the appliance of knowledge mining techniques to the realm of individual patient diagnosis.

2 Related Work

In today's digital world, medical data are being collected and are being made available to the medical researcher who has been interested in developing prediction models for diagnosis of disease. As a result, new research tool like knowledge discovery in databases (KDD), which incorporates data processing techniques, has become a preferred tool by medical researchers [3]. Heart disease may be a word that assigns to an enormous number of clinical conditions associated to heart. In this paper discussed about analyzing the varied data processing techniques introduced in recent years for heart condition prediction. Some of papers mentioned as follows:

Table 1. Summary of Different Methodology for Heart Disease Analysis

S. No	Research On	Methodology
1	Boot strap aggregation, multi-objective optimized voting for heart disease analysis[7]	Bagging with multi-objective weighted voting scheme
2	Heart disease prediction[8]	ANN algorithm
3	Heart disease prediction with minimum number of attributes [9]	Naïve Bayes classification by clustering and Decision Tree
4	Design and Implementing Heart Disease Prediction Using Naive Bayesian [10]	Navies Bayesian
5	Data Science and Its Application in Heart Disease Prediction [11]	Naïve Bayes, ANN, SVM
6	Prediction of Heart Disease Using Machine Learning [12]	Neural Networks
7	Heart Disease Prediction using Evolutionary Rule Learning [13]	Weighted Association Rule

The earlier discussed system which uses the different algorithms like KNN and other shows the integrated methodologies for designing the medical support system performance, i.e., 70.26% accurate. Many techniques which are discussed above used the analytical and mathematical approach with the data mining algorithms which leads to great performance and found effective in achieving the good results [5].

The proposed methodology is very useful in extracting the correct and necessary information and is modeled using the three modules which gains the performance improvement in terms of accuracy of 97%.The technique can be used in assembling in most of the samples and it is designed for each type of characteristic and properties as a set that have a direct featuring with each type heart diseases. There are different benchmarks set up while analyzing the condition of the patient which can significantly change the results for determining the actual cause of heart disease. And this approach is using the different types of the classifier for the classification of the characteristics of the disease along with the other constraints which are imposed while implementation of the design.

The approach used many classifiers named as Naïve Bayes, SVM, linear regression, instance-based learner, QDA (quadratic discriminate analysis) and obtained an accuracy of 84.16%.ELM model which used in generating the good type of the properties or characteristics which is then used as perfect attributes like age, sex, blood sugar, cholesterol, etc.

The approach studied can be applied to replace many high cost checkups or many other tests which are being so costly and can be avoided using the heterogeneous approach for the detection of the heart diseases. This method is used in real time problem where many patients' data have been gathered by medical practitioners [8]. The correctness shown by this approach is 80%. Models such as advanced algorithms, the systematic approaches have been applied to study and investigate the detection criteria based on the features [9]. Prior to that designing of the framework various different values have already been calculated to check the correctness of the approach proposed in the method. Disease prediction

classifier is very useful approach for detecting and prevention of the different types of heart diseases which come to be the useful approach for the analysis [10]. There is significant improvement in the detection of the disease can be found over the different types of other approaches which does not include these classifier in the framework and hence this method is more safe [11].

The research [14] on many aspects of decision making and producing the output in very low cost as compared to previously discussed methods for the detection of the diseases belonging to the heart and also focuses on heart disease diagnosis by considering previous useful information. Prediction of the various heart diseases can be achieved and greatly being improved. There may be some no risk factors concerning heart condition which needs to be considered while implementing the approach proposed. The probabilities of heart conditions in a patient and the subsequent characteristics have to be considered from the medical data of the patients which is very useful in generating the proper results, these include: age sex, BP, blood glucose, cholesterol, other physical characteristics of the patients which are very essential in making the system very useful for generating the good results. Various approaches studied this approach carefully and it is found very significant while analyzing and generating the correct detection results which [15] brings the concepts of hybrid model which can predict the disease and its severity within the less amount time and cost.

The most important aspect in designing and implementing the model involves the time and cost which needs to be addressed in the most of the research aspects and the solution of should be considerable for the all condition of the patients and must accepts any form of input for the universality and uniqueness.

Thus in order to make such things to work properly some modifications in the model has to be used and proposed. They are also various parameters and constraints which must also be addressed to work the system efficiently and effectively are the Accuracy, specificity, and sensitivity of the framework. The hybrid framework proposed has greater accuracy and required specificity and sensitivity as compared to the other previous approaches and hence must be applied in any form of heart the disease detection system.

Large amount of knowledge is produced and picked up by the healthcare organization on the day to day basis. To get intriguing knowledge, data innovation permits to extract the information through atomization of processes. Weighted Association Rule may be a sort of data processing technique wont to eliminate the manual task which also helps in extracting the information directly from the electronic records. This will help in decreasing the price of services and also helps in saving lives. Authors in [17] find the rule to predict patient's risk of having coronary disease. Test results have shown that vast majority of the rules helps in the best prediction of coronary illness. Theresa [18] proposed a model to precisely enhance the detection capabilities by improving the many aspects and is essentially very a great in building the new framework for the heart diseases. The accuracy has also been enhanced by using the new sort of classifier which is found working as per the characteristics of the system that is required for the proper utilization and working of the detection system [19]. The accuracy can also be improving many other aspects like the design and implementation specification which is supposed to improve the overall performance of the system for the detection and treatment of the fatal heart diseases [20]. Palaniappan & Awang [21] proposed a framework which is based on the completely new features and characteristics. This approach considers the various types of other classifiers which are quite easy to implement and allow some statistical data mining techniques. This may involve some mathematical modeling and which has to be done in producing the accurate output [22]. In [23], the approach mentioned uses the different advanced algorithms for the implementation of the design, this approach for designing and modeling includes the faster and easier detection based on the available data.

3 Proposed Technique

Considerable amount of research has already been done for predicting heart diseases much faster based on some rudimentary information available with a patient's history. These involve mostly ECG signals, cholesterol levels, blood pressure, sugar levels in the patient's blood etc. The patient's information is a result of data mining wherein the test results are stored and retrieved from the Hospital database when there is a need. The major challenge with the vast amount of information is the selection of the desired features. Hence in this proposed technique the features are selected in such a way that it may result in accurate predictions by the predictive algorithm.

Feature selection is a process where the framework automatically selects those features in your data that contribute most to the prediction variable or output in which you are interested. Having irrelevant features in the data can decrease the accuracy of many models, especially linear algorithms like linear and logistic regression. The "Cleveland heart disease dataset 2016" is available online from University of California, Irvine [23] going to use in the proposed technique. This proposed technique will select the most relevant features which may result with higher accuracy. These three may be the benefits of performing proposed technique for feature selection before modeling medical data will be:

- **Reduces Overfitting:** Less redundant data means lower chance to make opinions grounded on noise.
- **Improves Accuracy:** Less misleading data means modeling accuracy improves.
- **Reduces Training Time:** Less data means that algorithms train faster.

Determining the key features for the best model fitting machine learning is not an easy task. The main objective of this proposed technique is to accurately predict cardiovascular disease by comparison among different feature selection algorithms.

4 Conclusion

The approach discussed in this literature review is very significant in extracting the useful and crucial information and also gains the deepest knowledge from the huge amount of available information. This related work also summarized different aspects of different techniques and their comparisons based on classifiers for prediction and detection of different fatal Heart Diseases. Using advanced learning techniques to model and gain valuable information and novel insights. This study also discusses the number of different parameters such as accuracy and sensitivity and specificity for the prediction of different diseases.

Considering all above study here we propose feature selection technique which will lead to achieve more accuracy within effective time. More input attributes and specific characteristics like controllable and uncontrollable risk factors also help in generating better results.

References

- [1] Yanwei, X. et al. (2007). Combination data mining models with new medical data to predict outcome of coronary heart disease. In *Proceedings International Conference on Convergence Information Technology*, 868–872.

- [2] Abraham, R. et al. (2006). A comparative analysis of discretization methods for Medical Datamining with Naïve Bayesian classifier. In *Proceeding of IEEE International Conference on Information Technology*, 235 – 236.
- [3] Lavrac, N. (1999). Selected techniques for data mining in medicine. *Artificial Intelligence in Medical*, 16:3–23.
- [4] Rajathi, S. and Radhamani, G. (2016). Prediction and analysis of Rheumatic heart disease using KNN classification with ACO. *International Conference on Data Mining and Advanced Computing (SAPIENCE) IEEE*, 68-73.
- [5] Bashir, S. et al. (2016). HMV: A medical decision support framework using multi-layer classifiers for disease prediction. *Journal of Computational Science*, 13: 10-25.
- [6] Bialy, R. E., Salama, M. A. and Karam, O. (2016). An ensemble model for Heart disease data sets: a generalized model. In *Proceedings of the 10th International Conference on Informatics and Systems ACM*, 191- 196.
- [7] Bashir, S., Qamar, U. and Khan, F. H. (2015). BagMOOV: A novel ensemble for heart disease prediction bootstrap aggregation with multi-objective optimized voting. *Australasian Physical & Engineering Sciences in Medicine Springer*, 38: 305-323.
- [8] Ismaeel, A. S., Miri, A. and Chourishi, D. (2015). Using the extreme learning machine (elm) technique for heart disease diagnosis. In *Humanitarian Technology Conference (IHTC2015) IEEE Canada International*, 1-3.
- [9] Bashir, S., Qamar, U. and Javed, M. Y. (2014). An ensemble-based decision support framework for intelligent heart disease diagnosis. In *Information Society International Conference IEEE*, 259-264.
- [10] Bashir, S. et al. (2014). V5: a clinical decision support framework for heart disease prediction using majority vote- based classifier ensemble. *Arabian Journal for Science and Engineering Springer*, 39: 7771-7783.
- [11] Amin, S. U., Agarwal, K. and Beg, R. (2013). Genetic neural network-based data mining in prediction of heart disease using risk factors. *Information & Communication Technologies (ICT) IEEE*, 1227-1231.
- [12] Chen, A. H. et al. (2011). HDPS: Heart disease prediction system. *Computing in Cardiology IEEE*, 557-560.
- [13] Anbarasi, M. et al. (2010). Enhanced prediction of heart disease with feature subset selection using genetic algorithm. *International Journal of Engineering Science and Technology*, 2: 5370-5376.
- [14] Repaka, A. N., Ravikanti, S. D. and Franklin, R. G. (2019). Design and Implementing Heart Disease Prediction Using Naives Bayesian. In *3rd International Conference on Trends in Electronics and Informatics (ICOEI) India*, 292-297.
- [15] Junaid, M. J. A. and Kumar, R. (2020). Data Science and Its Application in Heart Disease Prediction. In *International Conference on Intelligent Engineering and Management (ICIEM)*, 396-400.
- [16] Gavhane, A. et al. (2018). Prediction of Heart Disease Using Machine Learning. In *Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 1275-1278.
- [17] Chauhan, A. et al. (2018). Heart Disease Prediction using Evolutionary Rule Learning. In *4th International Conference on Computational Intelligence & Communication Technology (CICT)*, 1-4.
- [18] Theresa, T., Princy, R. and Thomas, J. (2016). Human Heart Disease Prediction System using Data Mining Techniques. In *International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, 1-5.
- [19] Weimin, X., Sun, Y. and Lu, Y. (2006). Research and Application of Data Mining in Traditional Chinese Medical Clinic Diagnosis. In *Proceeding of IEEE 8th International Conference on Signal Processing*, 4: ISBN: 0-7803-9736-3.
- [20] Palaniappan, S. and Awang, R. (2008). Intelligent heart disease prediction system using data mining techniques. In *Proceeding of IEEE/ACS International Conference Doha*, 108-115.

- [21] Singh, J., Kamra, A. and Singh, H. (2016). Prediction of Heart Diseases Using Associative Classification, *IEEE*, 16:978- 1-5090-0893-3.
- [22] Pouriyeh, S. et al. (2017). A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease. In *22nd IEEE Symposium on Computers and Communication (ISCC 2017) Workshops - ICTS4eHealth*, 204-207.
- [23] Dua, D. and Graff, C. (2019). UCI Machine Learning Repository. In. *Irvine, CA: University of California, School of Information and Computer Science.*