

Prognostic System for Heart Disease using Machine Learning: A Review

¹Mahesh Kumar, ²K.Premkumar, ³R.Senthilkumar

^{1,2}U.G Student, ³Assistant Professor, Department of Computer Science Engineering,
Nehru Institute of Engineering and Technology Coimbatore.

¹maheshkumar0506@gmail.com, ²premkumars12g@gmail.com, ³sentinfophd@gmail.com

Abstract: In today's world it became difficult for daily routine check-up. The Heart disease system is an end user support and online consultation project. Here the motto behind it is to make a person to know about their heart related problem and according to it formulate them how much vital the disease is. It will be easy to access and keep track of their respective health. Thus, it's important to predict the disease as earliest. Attributes such as Bp, Cholesterol, Diabetes are fed into Classification methods of Machine Learning are been used to predict risk of heart disease.

Key words: HDPS, Machine Learning, Features Classification, Random forest algorithm, Dataset, Data-flow.

Introduction

“Heart Disease” refer to several types of heart condition which include disease like Heart Attack, Arrhythmia, Heart Failure. This may occur due to Blood Pressure, Cholesterol, Diabetes and many other factors. It makes difficult to identify the disease on the mentioned factors, by help of Machine Learning, Data Mining and Artificial Intelligence it becomes easy to recognize or catalog to find out disease among humans. It provides user interface to the end-user to enter the problem.

Heart Disease Prediction using Machine Learning consist of Train dataset as well as user's test dataset. There are various of methods of classification that work on the train and test dataset such as Decision Tree, Random forest, K Nearest Neighbour and etc. This application allows user to add their heart related issues. It's important to monitor heart disease because even a single high attribute can lead to various heart conditions.



Corresponding Author: R.Senthilkumar, Asst. Professor,
Department of CSE, Nehru Institute of Engineering and Technology
Coimbatore.
Mail: sentinfophd@gmail.com

The aim of Machine Learning into medical-care is to supplement user care for advance benefit. Machine Learning had made to assist different disease and give treatment properly. This Predictive analysis with different classifiers makes easy to improve in medical science.

Literature Survey

There are many works related to disease prediction systems which has been done using different machine learning algorithms and data mining techniques in medical centres.

[1]T John Peter, K Somasundaram It is using classification data mining techniques for prediction of heart disease. In this research paper, pattern recognition and data mining techniques are used into risk prediction models in the clinical domain where cardiovascular medicine is proposed. Using data mining technique, the data is then modeled and classified. Paper is also discovering set of rules based on given parameter about their health to predict the risk levels of patients. The performance is then calculated in measures of accuracy classification, rules generated, error classification and the results.

[2]Chaitrali Dangare, Sulabha Apte They are also using data mining classification techniques. They are using more number of input attributes like smoking and obesity to analyze and predict heart disease. Until now 13 attributes were taken in consideration for predicting the likelihood of patient whether he is having heart disease or not. Nowadays, Data mining techniques are playing vital role for predicting multiple diseases. Data mining techniques helps to reduce the number of tests. They have mainly concentrates on the diabetes, heart disease prediction and breast cancer etc.

[3]Nilakshi Waghulde, Nilima Patil Heart Disease Predication is proposed by Genetic neural approach. In this paper, it is mentioned how genetic algorithm and neural network works and how initialization of neural network is done using global optimization of genetic algorithm. It also calculates the number of hidden nodes to train the network with proper selection of architecture for neural network.

[4]Marjia Sultana , Mohammad Shorif Uddin heart disease is predicted by analyzing data mining techniques. Paper discusses the issue of prediction of heart disease on the basis of data mining techniques according to input attributes. Numbers of data sets are combined to measure the performance of various data mining techniques and based on that results are taken.

[5]J Thomas, R Theresa Princy Data mining techniques is used for Human heart disease prediction system. In this paper Input factors are considered for predicting the risk level of patient. The classification of patient risk level is done and then using data mining techniques the accuracy of risk level is calculated. The only limitation is time wasting and costly to hold a vast number of datasets with much occurring itemsets and less support or large itemsets.

[6]Mirpouya Mirmozaffari, Alireza Alinezhad proposed Data mining Apriori Algorithm for Heart Disease Prediction. In this paper, it's proposed that how the data mining is a solution which may be helpful in extracting the hidden pattern from medical dataset. This System was made of WEKA and MATLAB software within Apriori Algorithm. Accuracy obtained by this method is still not satisfied.

[7]**Saba Bashir, Khurram Bashir** proposed Improving Heart Disease Predication using Feature Selection approaches. In this paper, the motive was to focus on feature selection method and algorithms for doing experimental analysis and obtaining effective result. Also, to use of this in medical field using data science.

[8]**O.E. Taylor, Deedam-Okuchaba** proposed A Model to Detect Heart Disease using Machine Learning Algorithm. In this paper, dataset was being cleaned and processed making sure that there are no null values present in the dataset. Algorithms were also tested for accuracy. With help of correct algorithm, classification accuracy reached upto 88%.

[9]**Sankari Karthiga1 , M. Safish Mary2** proposed Early Prediction of Heart Disease Using Decision Tree Algorithm. In this paper, UCI repository dataset had been used for obtaining correct accuracy. From this it's cleared that SVM gives highest accuracy rate.

[10]**Chandrasegar Thirumala, Gautam Srivastava (Base Paper)** proposed Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques. They used different types of classifiers to study which algorithm gives more accurate result, with use of UCI dataset.

Related Work

We have collected and analyzed several years IEEE papers to get a refined visualization on Heart disease prediction system. There is a use of pattern recognition and data mining techniques into risk prediction models in the medical domain of cardiovascular medicine. The results obtained are then compared with the results of existing models, was found to be improved in further. The train dataset of heart disease patients was collected from the Kaggle which is used to discover patterns with NN. Feature Selection approaches is used for Improving Heart Disease Prediction [1-11]. The motive was to focus on feature selection method and algorithms for doing experimental analysis and obtaining effective result. Machine Learning classifier gives preferable outcomes, accuracy varying for different classifiers[12-24].

Proposed work:

The retrieval of the details of each patient is done using Decision Tree technique in Data Mining. And The performance of system is analyzed based on the accurate result prediction. The proposed system acts as a decision support system and will prove helpful for users as well as doctors in the diagnosis. A simple interface is made for the ease of user to understand and add their details. Patient's ID will be automatically be generated by the system. Data that was previously entered can be easily accessed. User can modify or update their details. After comparing patients data with input dataset, the results will be displayed. Later, feedback will be taken from the user, for any improvement of system. Based on the result of the patient, they can contact doctor for the further treatment if necessary.

Overall concept is about building a tree model (a random forest classifier) that will provide results with accuracy. The data is then inserted into model or classifier by which the probability of presence or absence of heart disease is predicted. Below there is entire process which is involved. There are combinations of 4 different databases in the dataset which is used for Heart disease Prediction.

There are total of 76 attributes in the database but all published experiments refer to using a subset of only 14 features. That's why, we have used the already processed UCI Cleveland dataset available in the Kaggle website as per convenience for analysis.

The description of all used 14 attributes is mentioned in below table.

S. No.	Attribute Description	Distinct Values of Attribute
1.	Age- represent the age of a person	Distinct values of all age groups
2.	Sex- detail about the gender of person (0- Female, 1-Male)	0,1
3.	CP- represents the severity of chest pain patient is suffering.	0,1,2,3
4.	Rest BP-It represents the patients BP.	Multiple values between 94& 200
5.	Chol-It shows the cholesterol level of the patient.	Multiple values between 126 & 564
6.	FBS-It represent the fasting blood sugar in the patient.	0,1
7.	Resting ECG-It shows the result of ECG	0,1,2
8.	Heartbeat- shows the max heartbeat of patient	Multiple values from 71 to 202
9.	Exang- used to identify if there is an exercise induced angina. If yes=1 or else no=0	0,1

Conclusion

Heart Disease tends to be fatal natured. This disease creates life threatening complexities such as heart attack which eventually leads to death. With the increasing cases of deaths due to heart diseases, it has become mandatory to develop a system to predict heart diseases effectively and accurately. Various researches were carried out for the same by many people using effective techniques. Although various classification techniques were used for predicting heart disease, here we've chosen random forest since because of its coherence and accuracy. Different attribute selection measures like Information Gain, Gain Ratio, Gini Index and Distance measure can be used by random forest algorithm.

References

1. Heart Disease Dataset - <https://www.kaggle.com/c/heart-disease> dated: Sept 2018
2. Effective heart disease prediction using hybrid machine learning techniques Senthilkumar Mohan, ChandrasegarThirumalai, Gautam Srivastava IEEE Access 7, 81542-81554, 2019
3. Heart Disease Prediction Using Machine Learning Algorithms A Singh, R Kumar - 2020 International Conference, 2020-ieeexplore.ieee.org
4. Heart disease prediction using machine learning techniques: a survey by VV Ramalingam, Ayantan Dandapath, M Karthik Raja International Journal of Engineering and Technology 7(2.8), 684-687, 2018
5. Prediction of heart disease using machine learning Aditi Gavhane, GouthamiKokkula, Isha Pandya, Kailas Devadkar 2018
6. Karthick, R., et al. "Overcome the challenges in bio-medical instruments using IOT–A review." *Materials Today: Proceedings* (2020). <https://doi.org/10.1016/j.matpr.2020.08.420>
7. Karthick, R., et al. "A Geographical Review: Novel Coronavirus (COVID-19) Pandemic." *A Geographical Review: Novel Coronavirus (COVID-19) Pandemic* (October 16, 2020). *Asian Journal of Applied Science and Technology (AJAST)(Quarterly International Journal)* Volume 4 (2020): 44-50.
8. Sathiyathan, N. "Medical Image Compression Using View Compensated Wavelet Transform." *Journal of Global Research in Computer Science* 9.9 (2018): 01-04.
9. Karthick, R., and M. Sundararajan. "SPIDER-based out-of-order execution scheme for Ht-MPSOC." *International Journal of Advanced Intelligence paradigms* 19.1 (2021): 28-41. <https://doi.org/10.1504/IJAIP.2021.114581>
10. Sabarish, P., et al. "An Energy Efficient Microwave Based Wireless Solar Power Transmission System." *IOP Conference Series: Materials Science and Engineering*. Vol. 937. No. 1. IOP Publishing, 2020. doi:10.1088/1757-899X/937/1/012013
11. Vijayalakshmi, S., et al. "Implementation of a new Bi-Directional Switch multilevel Inverter for the reduction of harmonics." *IOP Conference Series: Materials Science and Engineering*. Vol. 937. No. 1. IOP Publishing, 2020. doi:10.1088/1757-899X/937/1/012026
12. Karthick, R., and M. Sundararajan. "Hardware Evaluation of Second Round SHA-3 Candidates Using FPGA (April 2, 2014)." *International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014)* 2.2.
13. Karthick, R., et al. "High resolution image scaling using fuzzy based FPGA implementation." *Asian Journal of Applied Science and Technology (AJAST)* 3.1 (2019): 215-221.
14. P. Sabarish, R. Karthick, A. Sindhu, N. Sathiyathan, Investigation on performance of solar photovoltaic fed hybrid semi impedance source converters, *Materials Today: Proceedings*, 2020, <https://doi.org/10.1016/j.matpr.2020.08.390>

15. Karthick, R., A. Manoj Prabakaran, and P. Selvaprasanth. "Internet of things based high security border surveillance strategy." *Asian Journal of Applied Science and Technology (AJAST)* Volume 3 (2019): 94-100.
16. Karthick, R., and M. Sundararajan. "A novel 3-D-IC test architecture-a review." *International Journal of Engineering and Technology (UAE)* 7.1.1 (2018): 582-586.
17. Karthick, R., and M. Sundararajan. "Design and implementation of low power testing using advanced razor based processor." *International Journal of Applied Engineering Research* 12.17 (2017): 6384-6390.
18. Karthick, R., and M. Sundararajan. "A Reconfigurable Method for TimeCorrelatedMimo Channels with a Decision Feedback Receiver." *International Journal of Applied Engineering Research* 12.15 (2017): 5234-5241.
19. Karthick, R., and M. Sundararajan. "PSO based out-of-order (ooo) execution scheme for HT-MPSOC." *Journal of Advanced Research in Dynamical and Control Systems* 9 (2017): 1969.
20. Karthick, R. "Deep Learning For Age Group Classification System." *International Journal Of Advances In Signal And Image Sciences* 4.2 (2018): 16-22.
21. Karthick, R., and P. Meenalochini. "Implementation of data cache block (DCB) in shared processor using field-programmable gate array (FPGA)." *Journal of the National Science Foundation of Sri Lanka* 48.4 (2020). <http://doi.org/10.4038/jnsfsr.v48i4.10340>
22. Suresh, Helina Rajini, et al. "Suppression of four wave mixing effect in DWDM system." *Materials Today: Proceedings* (2021). <https://doi.org/10.1016/j.matpr.2020.11.545>
23. M. Sheik Dawood, S. Sakena Benazer, N. Nanthini, R. Devika, R. Karthick, Design of rectenna for wireless sensor networks, *Materials Today: Proceedings*, 2021. <https://doi.org/10.1016/j.matpr.2020.11.905>
24. M. Sheik Dawood, S. Sakena Benazer, R. Karthick, R. Senthil Ganesh, S. Sugirtha Mary, Performance analysis of efficient video transmission using EvalSVC, EvalVid-NT, EvalVid, *Materials Today: Proceedings*, 2021. <https://doi.org/10.1016/j.matpr.2021.02.287>