Journal of Critical Reviews

ISSN- 2394-5125 Vol 7, Issue 12, 2020

A COMPREHENSIVE REVIEW OF HEART DISEASE PREDICTION USING MACHINE LEARNING

Dr Dilbag Singh [1], Jasjit Singh Samagh[2]

[1] Professor, Department of CSA, Chaudhary Devi Lal University, Sirsa, India.
[2]Research Scholar, Department of CSA, Chaudhary Devi Lal University, Sirsa, India
Email ID:jasjitsamagh@gmail.com

Received: 07.03.2020 *Revised: 09.04.2020* *Accepted: 10.05.2020*

Abstract
Heart disease is the most common causes of death in underdeveloped, developing and even the developed countries, lacs of the people die every year due to this. Early detection and accurate prediction of this disease may certainly reduce the mortality rate. For the detection and prediction of cardiovascular diseases efficiently, there is a need of health history and causes which lead to heart diseases. Machine Learning plays a vital role in predicting and maintaining essential data about the Heart disease using its various machine learning techniques and Hybrid models. In present study the review of various research works carried out in respect to application of machine learning in prediction of heart diseases on the basis of datasets is presented. Present research work will assist the medical practitioners in predicting the heart threats well in time so as to take the measures to control the mishap.

Keywords Heart disease, Prediction, Machine learning.

DOI: http://dx.doi.org/10.31838/jcr.07.12.54

INTRODUCTION

Cardiovascular disease relates to contraction or blockage of blood vessels. Debilitating of heart muscle, valve or beat additionally are viewed as type of heart disease. Coronary illness is silent killers as their symptoms are very difficult to detect. There is a rapid increase in heart disease and as per World Health Organization calculations over 17 million people die every year due to heart disease like heart attacks and strokes, therefore, it is necessary to have a record of crucial evidence and health problems leading to Heart disease. Heart disease of newly born babies is called as congenital Heart Disease and at the later ages it is called acquired Heart Disease. Number of tests are performed namely blood pressure, ECG, auscultation, cholesterol and blood sugar for heart disease diagnosis. These test procedures takes a long time and administering of medicines gets delayed which sometimes may have adverse effect on the patient. Machine learning may help doctors, pathologists to reduce the timing of such tests and results will be more accurate because amount of data is increased. With machine learning it is very easy to get knowledge from a large data which is not possible for man to analyse.

MACHINE LEARNING: AN OUTLINE

In 1959, the term Machine learning was given by Arthur Samuel which defines Machine learning as a discipline which gives computers the power to discover with not being programmed explicitly [1]

Machine learning is the process to generate the algorithms based on the previous inputs and past experiences. [1]

To understand the use of machine learning more accurately we will consider some of the examples where the concept machine learning is applied. Such examples are self-driving car, Google listening search, dashboards on Amazon, showcases on Netflix and many more are the examples of machine learning. [1]

Let us understand machine learning with one more example now just imagine that you are using a news application, but the application is not able to provide you with latest notifications and updates on time. And now this will make you think that the application is of no use and it is not making any aid to me. So, in this case the concept of machine learning is used through which we can achieve the best from the application or software. Machine learning uses the concept that enables the automatic notification feature provider by the application. This is a very common real-life example of why we need the concept like machine learning. [1]

MACHINE LEARNING APPROACHES

For solving a problem basically four types of Machine Learning approaches being used they are supervised, unsupervised, semi supervised and reinforcement learning.

Supervised Learning

Supervised learning is the machine learning approach that consists of techniques that learn from externally supplied examples that results in a general theory which makes prediction about future instances. An outcome or output-variable guides the whole learning process. A learning process assesses an unknown sample from known sample, where sample cold be both input or output and where output is labelled as classification or regression. The process employs a known dataset called training dataset that enables predictions. A training dataset consists of input data and response values. Learning process seeks to build a model from the dataset that can make predictions of the response values for a new dataset. A test dataset is often used to confirm the performance of a Machine Learning model. Larger training datasets often produces the models with greater predictive power that can better simplify new datasets. Examples of supervised learning algorithms are Decision tree, KNN, SVM, ANN, and Random Forest etc. [2]

Unsupervised Learning

Unsupervised learning is a type of machine learning approach which is used to draw conclusions from datasets which consist of input data with no labelled responses. In this approach there are basically two learning tasks namely Association and clustering. Apriori is considered as the best algorithm used in association rule. K-means clustering and association rule learning algorithm are the commonly used algorithms in unsupervised learning approach. [3]

Semi-supervised Learning
Semi-supervised learning is a machine learning approach which is mainly the combination of supervised and unsupervised learning in which portion of data is partially labelled. This learning approach is mostly practiced in speech recognition, in classification of webpages and in genetic sequencing. [3]

Reinforcement Learning
Reinforcement Learning is a machine learning strategy which manages with how software agents will be utilized consequently to decide the ideal action of a particular setting in order to boost its performance and also sends the feedback for the specialists to learn its behaviour. Classification and control are the learning tasks associated with Reinforcement learning. Self-driving car is one of the application where we have seen the use of reinforcement learning. Q-learning, Temporal difference and the Deep Adversarial Networks are the commonly used Reinforcement learning algorithms. [1]

ML CLASSIFICATION TECHNIQES
The Machine Learning classification techniques are as follows

Regression
Regression is one of the supervised learning model which aims to deliver the prediction of an
Output variable in accordance with its input and which are later identified. Mostly recognized algorithms are linear regression, logistic regression and stepwise regression. There are complex regression algorithms developed like ordinary least squares regression and multivariate adaptive regression [4].

Clustering
Clustering is an unsupervised learning model basically used in form of clusters to discover natural groupings of data. Commonly used clustering techniques are the k-means technique, the hierarchical technique and the expectation maximization technique [4].

Bayesian Models
Bayesian models uses supervised learning for solving either classification or regression problems. Most commonly used algorithms are naive bayes, gaussian naive bayes, multinomial naive bayes, bayesian network, mixture of gaussians, and bayesian belief network [4].

K- Nearest Neighbor
K-nearest neighbor (KNN) technique is simple but powerful classification algorithm in which the value of K helps in measuring the nearest neighbor as it clarifies how many nearest neighbors will be needed to examine as the class of a sample data point will be described through it. This technique is further classified into structure based KNN and structure less KNN. In structure based technique the basic structure of the data with less mechanism is used and it includes training data samples whereas in structure less technique entire data is divided into sample data and training data points and to find the nearest neighbor the distance is calculated between the them and the point that have the smallest distance is the nearest neighbor [4].

Support Vector Machines
Support vector machines (SVMs) were introduced first to be used in statistical learning theory. SVM is basically a binary classifier that create a linear separating hyper plane to sort data position. SVMs are basically used in classification, regression, and clustering. In case of global optimization SVMs deal with more complex problems which appear in high dimensional spaces which makes them attractive in various applications. Commonly used SVM algorithms are the support vector regression, least squares support vector machine, and successive projection algorithm-support vector machine [4].

RELATED WORK
Youness Khourdifi et.al. [5] compared the algorithms with different performance measures with the help of machine learning. For prediction algorithms worked differently.In this study the best results are shown by K-Nearest Neighbour, and Random Forest and Artificial Neural Network. Then he merged the algorithms and try to check the performance so as to check if it will be more efficient or not and later result is applied to heart disease data set and his proposed models resulted in better accuracy of 99.65%

Youness Khourdifi et.al [6] evaluated the result of proposed model and compared with the algorithms having different performance measures which resulted into better performance and effectiveness as compared to K-Nearest Neighbour, Random Forest, Naïve Bayes, Support Vector Machine SVM and Artificial Neural Network.

Komal Saini et.al. [7] Proposed an IoT framework for the prediction of the heart disease. Wearable sensing devices which are used to collect the data from the body of the patients and then storage and processing of data using communication standards. Among the communication standards which transfer the data collected to the storage devices, Bluetooth (BLE) has been well suited for the health care application and for the storage of the huge amount of data generated by the wearable sensor cloud storage and big data management is considered as the best. Finally, for the analysis of the data to predict a model data miming tools and machine learning algorithms are used to get the accurate results but the accuracy of all the algorithms vary according to the parameters and data collected.

SenthilKumar Mohan. et.al. [8] proposed the hybrid approach and named it as hybrid random forest with linear model (HRFLM). It is the combination of Random forest and linear model in which basic characteristics of both the algorithms were combined. The novel approach resulted quite accurately in the prediction of heart disease. Accuracy achieved was 88.7%.

Pooja et.al. [9] provided different ways to implement the concept of machine learning and to get the best results from out there. Basically, machine learning is the concept to analyse the huge amount of data. In past centuries, the maximum amount of time was let's say wasted by our programmers in order to analyse the huge volume of data on their own and getting the result out of it. We all know that it took hours and hours sometimes even years to estimate the data in bulk. When at last the time of results came, then again, they get errors in achieving the desired output. So, this was a very poor and bad approach as it consumes more time and efforts and also didn't produced the results according to the desire of the user.

COMPARATIVE ANALYSIS
The table below presents the work carried out on machine learning for prediction of heart diseases. In the table 1 the work carried out by different researchers is presented along with the prediction accuracy. It will assist the medical practitioners in determining the better approach of machine learning for effective predication of heart disease threats.

Table-1: Performance comparison of Machine Learning techniques in heart disease prediction

YEAR	AUTHOR'S NAME	DESCRIPTION	ML TECHNIQUES USED	ACCURACY
2012	G. Parthiban et al.[10]	Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients	Naïve Bayes and Support Vector Machines	Naïve Bayes resulted in 74% of accuracy whereas SVM have proven to be classified technique with excellent predictive performance of 94.60%
2012	Vikas Chaurasia et al.[11]	Data Mining Approach to Detect Heart Diseases	Naïve Bayes,J48, Decision tree and Bagging 10 Fold cross validation method	Bagging algorithm give better accuracy of 85.03% with total time of .05seconds to build model in diagnosis of Heart Disease
2015	Boshra Brahmi et al.[12]	Prediction and Diagnosis of Heart Disease by Data Mining Techniques	K-Nearest Neighbor, SMO J48 and Naïve Bayes	It was seen that J48 performed better results than the other techniques.
2015	K.Vembandasamy et al.[13]	Heart Diseases Detection Using Naive Bayes Algorithm	Naïve Bayes	Naïve Bayes algorithm provides 86.4198% of accuracy with minimum time
2015	Ahmed Fawzi Otoom et al.[14]	Effective Diagnosis and Monitoring of Heart Disease	Build an intelligent classifier using ML algorithms (Naïve Bayes, SVM and FT)	Gives accuracy of more than 85% with cross validation test moreover the monitoring algorithm provides a 100% detection rate
2016	S. Seema et al[15]	Chronic Disease Prediction by mining the data.	SVM, Decision Tree and Naïve Bayes	SVM achieved the highest accuracy then the other techniques in case of heart disease which was 95.56% and in case of Diabetes Highest accuracy of 73.588% achieved by Naïve Bayes.
2016	K. Gomathi et al.[16]	Multi Disease Prediction using Data Mining Techniques.	J48 and Naïve Bayes	Using Naïve Bayes algorithm the accuracy for Heart Disease prediction is 79%, for Diabetes is 77.6% and for Breast Cancer is 82.5% UsingJ48 algorithm the accuracy for Heart Disease prediction is 77%,for Diabetes is 100% and for Breast Cancer is 75.5%
2016	Ashwini Shetty A et al.[17]	Different Data Mining Approaches for Predicting Heart Disease	Neural Network Hybrid System	In case of Neural Network accuracy is predicted to be 84% For this system the accuracy resulted is 89%
2016	Ayon Dey et al.[18]	Analysis of Supervised Machine Learning Algorithms for Heart Disease Prediction with Reduced Number of Attributes using PCA	SVM, Naive Bayes and Decision tree without PCA SVM, Naive Bayes and Decision tree with PCA	Decision Tree has better performance SVM outperforms the other two

2017	Aized Amin Soofi et al.[19]	Classification Techniques in Machine Learning: Applications and Issues	Bayesian Network, Decision Tree induction, KNN classifier and SVM	It was hard to conclude which technique was better qua others
2017	Min Chen et al.[20]	Disease Prediction by Machine Learning Over Big Data from Healthcare Communities	Convolutional neural network (CNN)-based multimodal disease risk prediction algorithm	It resulted in accuracy of 94.8% with a convergence speed
2018	Kusuma et.al. [21]	Machine Learning and Deep Learning Methods in Heart Disease	Deep Learning Technique and tools	Accuracy achieved by this is 60%
			Machine learning Technique and SVM	Accuracy achieved is 30%
2018	Shadman Nashif et.al. [22]	Heart Disease Detection by Using Machine Learning Algorithms and a Real Time Cardiovascular Health Monitoring System	Machine Learning Techniques with Java based open access data Mining platform and WEKA	SVM algorithm resulted in 97.53% accuracy. Sensitivity and specificity of algorithm noted at 97.50% and 94.94%
2019	Senthilkumar Mohan et al.[8]	Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques	Hybrid random forest with Linear Model (HRFLM)	Accuracy achieved is 88.7% with this random model
2019	Youness Khourdifi et al.[6]	Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization	Hybrid Model combining FCBF, PSO and ACO	Accuracy achieved by this model is 99.65%
2019	Youness Khourdifi et al.[5]	The Hybrid Machine Learning Model Based on Random Forest Optimized by PSO and ACO for Predicting Heart	After Optimization PA-RF	Accuracy achieved is 99.6%
			Before Optimization: -	Accuracy achieved was as under
			Support Vector Machine	83.6%
			Naïve Bayes	83.6%
			Random Forest	81.35%

From the above table, it has become clear that there are various attribute agents which can enhance the performance of Machine learning techniques for predicting the Heart Disease.

These are as follows:-

- Ensemble: The approach of combining different technique with diverse strengths is known as Ensemble. There are various machine learning techniques used for heart disease prediction. Each technique has its own strength and weakness. Ensemble method combine the weaker learners to create stronger ones which helps in averaging out biases of different underlying techniques and also reducing the inconsistency in the prediction of heart diseases.
  Research papers [5] [6] [8] [17] [20] used the ensemble approach in model building for predicting the heart disease. It has become clear from the said papers that ensemble approach enhances the performance of heart disease prediction and accuracy is also recorded more than 90% which is very less in comparison with model using simple Machine learning techniques.
- PCA: Principal Component Analysis helps in finding the most effective transformation of existing attributes through a linear transformation technique. It helps with dimensionality reduction, which makes thing faster by reducing the size of dataset to be stored and processed which will increase the performance in predicting the heart disease.

Paper [18] clarify the use of Principal Component analysis. PCA helps in reducing the number of attributes which reduces the size of the dataset, some algorithms outperform the other algorithms, hence resulting in enhancing the performance for predicting the heart disease.

- Feature Selection: In the heart disease datasets, the number of features can reach up to tens of thousands. However for the better prediction of heart diseases about 14 attributes are sufficient as a large number of irrelevant and redundant attributes usually give incorrect results, high cost and consume more time. Lesser the attributes better would be the performance in predicting the heart disease.
  In Paper [6], Fast Correlation Based Feature Selection method is used. To enhance the quality of heart disease the redundant features were filtered using feature selection method. The accuracy of 99.6% is achieved using feature selection method.
- Optimizers: Optimization algorithms are used to find the best solution of the problem by minimization or maximization the objective function without violating resource constraints. Optimization algorithms can be applied to improve the performance of classifiers by tuning parameters so as to enhance the performance in heart disease prediction.
  Paper [5] [6] used Particle Swarm Optimization and Ant Colony Optimization algorithms. It is seen that with the use

of optimization algorithms in the hybrid approach the accuracy for heart disease prediction achieved is 99.65%

- Tools: There are many data mining tools used in machine learning for prediction of Heart disease such as Weka, Matlab, SaS etc. It is seen that Weka is mostly used because it is meant only for machine learning and it has inbuilt machine learning algorithms which makes data mining easier hence resulted in high performance then the other data mining tools in predicting the heart disease.
  Finding of paper [22] clearly reveals that the use of WEKA tool, the prediction of the heart disease resulted very fast and accuracy achieved to 97.5%

CONCLUSION

This paper presents a systematic review of different types of machine learning techniques to predict heart diseases. An analytical study has been conducted for the existing techniques and compared for finding out the efficient and accurate systems. An exploration has been left for designing ensemble models using Feature Selection method and Optimization techniques. These novel and innovative models may improve the performance of heart diseases prediction, with the help of which early diagnosis of heart diseases in patients can be identified. These approaches may help the medical practitioners in preventing the mishaps by giving preventive treatment.

REFERENCES

1. Pooja, Aakansha Sharma and Ankush Sharma (2018). Machine Learning: A Review of Techniques of Machine Learning, Journal of Applied Science and Computations Volume 5, Issue 7, July /2018.ISSN NO: 1076-5131.
2. Senthilkumar Mohan, Chandrasegar Thirumalai, and Gautam Srivasatava (2019). Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques, IEEE access volume7, 2019.
3. Shadman Nashif, Md. Rakib Raihan, Md. Rasedul Islam,and Mohammad Hasan Imam (2018).Heart Disease Detection by Using Machine Learning Algorithms and a Real Time Cardiovascular Health Monitoring System, World Journal of Engineering and Technology, 2018, 6, 854-873.
4. Konstantinos G. Liakos, Patrizia Busato, Dimitrios Moshou , Simon Pearson and Dionysis Bochtis (2018).Machine Learning in Agriculture: A Review,Sensors 2018, 18, 2674.
5. Youness Khourdifi and Mohamed Bahaji (2019). Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization, International Journal of Intelligent Engineering and Systems, Vol.12, No.1, 2019.
6. Youness Khourdifi and Mohamed Bahaji (2019). The Hybrid Machine Learning Model Based on Random Forest Optimized by PSO and ACO for Predicting Heart, ICCWCS 2019, April 24-25, Kenitra, Morocco.
7. Komal Saini and Sandeep Sharma (2019). Review on the Heart Disease Detection Using IoT Framework, International Journal of Computer Sciences and Engineering Vol.7(3), Mar 2019, E-ISSN: 2347-2693.
8. Senthilkumar Mohan, Chandrasegar Thirumalai, and Gautam Srivasatava (2019). Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques, IEEE access volume7, 2019.
9. Pooja, Aakansha Sharma and Ankush Sharma (2018). Machine Learning: A Review of Techniques of Machine Learning, Journal of Applied Science and Computations Volume 5, Issue 7, July /2018.ISSN NO: 1076-5131.
10. G. Parthiban and S.K.Srivatsa (2012). Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients, International Journal of Applied Information Systems (IJAIS) – ISSN: 2249-0868.
11. Vikas Chaurasia and Saurabh Pal (2012). Data Mining Approach to Detect Heart Diseases, International Journal of Advanced Computer Science and Information Technology (IJACSIT)Vol. 2, No. 4, Month Year, Page: 56-66, ISSN: 2296-1739.
12. Boshra Brahmi, Mirsaeid Hosseini Shirvani (2015). Prediction and Diagnosis of Heart Disease by Data Mining Techniques, Journals of Multidisciplinary Engineering Science and Technology, vol.2, 2 February 2015, pp.164-168.
13. K.Vembandasamy, R.Sasipriyaand E.Deepa (2015). Heart Diseases Detection Using Naive Bayes Algorithm, International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 9, September 2015.
14. Ahmed Fawzi Otoom, Emad E. Abdallah, Yousef Kilani, Ahmed Kefaye and Mohammad Ashour (2015). Effective Diagnosis and Monitoring of Heart Disease, International Journal of Software Engineering and Its Applications Vol. 9, No. 1 (2015), pp. 143-156.
15. Dr.S.Seema Shedole, Kumari Deepika (2016). Predictive analytics to prevent and control chronic disease, https://www.researchgate.net/punlication/316530782, January 2016.
16. K.Gomathi, Dr.D.Shanmuga Priyaa (2016). Multi Disease Prediction using Data Mining Techniques, International Journal of System and Software Engineering, December 2016, pp.12-14.
17. Ashwini Shetty A, Chandra Naik (2016). Different Data Mining Approaches for Predicting Heart Disease, International Journal of Innovative in Science Engineering and Technology, Vol.5, May 2016, pp.277- 281.
18. Ayon Dey, Jyoti Singh and Neeta Singh (2016). Analysis of Supervised Machine Learning Algorithms for Heart Disease Prediction with Reduced Number of Attributes using Principal Component Analysis), International Journal of Computer Applications (0975 – 8887) Volume 140 – No.2, April 2016.
19. Aized Amin Soofi and Arshad Awan (2017). Classification Techniques in Machine Learning: Applications and Issues, Journal of Basic & Applied Sciences, 2017, 13, 459-465.
20. Min Chen, Yixue Hao, Kai Hwang, Lu Wang, and Lin Wang (2017). Disease Prediction by Machine Learning Over Big Data from Healthcare Communities, IEEE Access.
21. Kusuma and Divya Udayan (2018). Machine Learning and Deep Learning Methods in Heart Disease (HD) Research, International Journal of Pure and Applied Mathematics Volume 119 No. 18 2018, 1483-1496.
22. Shadman Nashif, Md. Rakib Raihan, Md. Rasedul Islam,and Mohammad Hasan Imam (2018).Heart Disease Detection by Using Machine Learning Algorithms and a Real Time Cardiovascular Health Monitoring System, World Journal of Engineering and Technology, 2018, 6, 854-873.