EasyChair Preprint

№ 7969

A Survey on Machine Learning Algorithms for Predicting Cardiovascular Disease

R Sathya, J Shanthini and S Karthik

May 21, 2022

A Survey on Machine Learning Algorithms for Predicting Cardiovascular Disease

Sathya R[1], Dr. Shanthini.J[2], Dr. Karthik.S[3]

[1]Full Time PhD Scholar, SNS College of Technology, Coimbatore, TamilNadu, India
[2]Professor/CSE, Dr. NGP Institute of Technology, Coimbatore, TamilNadu, India
[3]Professor/Dean, CSE, SNS College of Technology, Coimbatore, TamilNadu, India
sathyagbsri@gmail.com

Abstract. Cardiovascular disease(CVD) is a kind of disease that distress the heart or blood vessels (veins and arteries). It can be initiated by means of a mixture of socio-economic, behavioral, and ecological risk features, including high blood pressure, unhealthy diet, high cholesterol, diabetes, tobacco use, physical inactivity, harmful use of alcohol and stress. Family history, traditional background, sex, and age can also distress a person's threat of cardiovascular disease. Machine learning algorithms can be used to predict the heart diseases more accurately and precisely. And the risk factors for heart disease can also be predicted effectively using Machine learning techniques and algorithms. This paper suggests a number of models built on algorithms and methods, to examine their performance. Best frequently used Supervised ML algorithms by the scholars are SVM, KNN, Decision Trees, Random Forest, Ensemble models.

Keywords: Machine Learning(ML), SMOTE, Neural Network, Random Forest, Feed forward backpropagation, Waikato Environment for Knowledge Analysis(WEKA).

1 Introduction

The heart is a strong organ that pushes blood all over the body through the circulatory system and the system contains a heart, blood vessels and blood. The blood pumped from the heart is given to the tissues and organs with oxygen and nutrients through the blood vessels, it also carries carbon dioxide and it is considered as a metabolic waste.

Globally, nearly an estimated 18.6 million people died because of cardiovascular disease in 2019, which represents 32% of all global deaths. Out of these, 85% were due to stroke and heart attack. About three quarters of heart disease deaths occurs in small and medium revenue countries.

Medical organization in and around the world have collected the data related to health issues. All data collected can be used by different ML techniques to increase some valuable visions. The parameters used for collecting health related data are very huge and can be noisy too. These datasets are too vast to be handled by the humans and so can be simply sightseen by various ML practices. All the practices are right beneficial to forecast the occurrence or nonexistence of heart associated diseases more precisely and correctly.

2 Related Works

By using Machine learning algorithms, a lot of survey and studies has been on the prediction of cardiovascular disease.

Primarily, Shaikh Abdul Hannan et. al. [1] used SVM (Support Vector Machine) and FFBP (Feed Forward Back Propagation) method for the prediction of cardiovascular disease. SVM works based on Structural Risk Minimization technique and it's a novel machine learning technique. The Feed Forward Back Propagation consists of 3 layers: i) An input layer, ii) An output layer and iii) A thin hidden layer.

The time taken by SVM algorithm is 5-10 sec and FFBP algorithm is 45-50mins. It provides less suitable outcomes for medical treatment.

Poornima Singh et al. [2] developed a prediction system(EHDPS) for finding heart disease effectively by using neural network for predicting the risk level of heart disease. In artificial neural network, Multilayer Perceptron Neural Network(MLPNN) is one of the finest models. MLPNN consist of i) one input layer, ii) one or more hidden layers and iii) one output layer. The input nodes give the values to the first hidden layer, and then to the second and so on till the system produces results. Totally 303 records are there in thhe dataset and again it was separated into two parts: i) Training set with 40% data and ii) Testing set with 60% data. The system predicts heart disease with 100% accuracy as the neural network was applied on training dataset. The result shows that there are zero False Negative and Zero False Positive entries.

Mr. J. Santhana Krishnan and Dr.S.Geetha et al. [3] proposed a method that predicts the Male patient affected by heart disease using classification techniques. Different factors which influences the heart disease such as Evidences, Common types and their threat features has been clearly enlightened. WEKA is a Data Mining tool used for predicting the heart disease with three different interfaces: i) Naive Bayes, ii) Artificial Neural Networks and iii) Decision Tree. With the help of these techniques, it can be easily predicted with the datasets. CART, C4.5, CHAID, J48, ID3 Algorithms are the various Decision tree techniques. For prediction of heart disease, Naïve Bayes is used.

Muhammad Waqar et al. [4] proposed a Synthetic Minority Oversampling Technique (SMOTE). The paper presents a approach to foretell the heart disease with the specified imbalance data. SMOTE techniques is used to handle the imbalance dataset because it consists of positive and negative classes. It removes the necessary of feature engineering for the organization of specified dataset. The proposed neural network model classifies the dataset without any data preprocessing and it takes less time to produce the reliable results.

Prashasti Kanikar et al. [5] used Support Vector Machines and Bayesian Classification for predicting the cardiovascular disease. SVM method is easy to understand outcomes and easy error detection. Naïve Bayes algorithm delivers minimum error rate and works with small volume of data. It determines that Support Vector Machine algorithm is enhanced than Naïve Bayes algorithm as Support Vector Machine works better for large datasets.

Avinash Golande et al. [6] suggested Heart Disease Prediction with effective ML Methods in which Experts apply a little data mining approaches that are accessible to sustenance the experts or specialists to discriminate the disease. Different approaches and strategies are used. The results show the accuracy of 86.60% in finding the heart disease by using Decision tree(DT), K – Nearest and Naïve Bayes.

Senthil Kumar Mohan et al. [7] suggested a prediction system for heart disease using Hybrid ML Techniques in which the critical aspects can be found by using ML methodologies and also to rise the accuracy in the expectation of cardiovascular malady. It provides a precision level of 88.7% with the help of Hybrid Random Forest with a Liner Model(HRFLM).

A.S.Abdullah and R.R.Rajalaxmi et al.[8] used a method to sort the various metabolic syndromes and diseases by using identification and prediction in DM. To measure the actions associated to Coronary Heart disease, they used the Decision Tree classification algorithm. Random Forest Classifier is used to foresee the existence of several actions interrelated to the patient histories, risk issue preventions, and to increase the precision of the complete prediction.

A.H.Alkeshuosh, M.Z.Moghadam, I.Al Mansoori, and M.Abdar et al. [9] suggested Classification Rule Mining is the best task in Data Mining. To store a very large amount of data in databases, PSO-based algorithms for classification rule mining are used. Decision tree-based C4.5 algorithm and PSO-based algorithms are compared to produce good accuracy results. The algorithm proposes good rules in heart disease prediction.

H.A.Esfahani and M.Ghazanfari et al. [10] proposed a method to diagnose heart disease in earlier. The dataset with 303 patient records is taken and analyzed by using three classifiers including neural network, Naïve Bayes, and Rough set. They have used a weighted majority vote. The ensemble classifier gives the fusion outputs with 86.8% of F-Measure.

Table 1. Comparison of techniques

S.No	Research (year)	Technique	Accuracy
1.	A.S.Abdullah, R.R.Rajalaxmi et al(2017)	Random Forest Classifier	63.33%
2.	A.H.Alkeshuosh, M.Z.Moghadam, I.Al Mansoori, M.Abdaret al.(2017)	Particle Swarm Optimization	Better results and rules.
3.	H.A.Esfahani, M.Ghazanfari et al.(2017)	Ensemble Classifier	86.8%
4.	Shaikh Abdul Hannan et al. (2018)	SVM Feedforward Backpropagation NN	50% to 60% output
5.	Poornima Singh et al. (2018)	Neural Network	100%
6.	Mr.J.Santhana Krishnan, Dr.S Geetha (2019)	Naïve Bayes, Decision Tree	91%
7.	Muhammad Waqar et al(2021)	SMOTE-based artificial neural network	100%

8.	Prashasti Kanikar et al. (2016)	SVM (RBF) Naïve Bayes	57% 52%
9.	Avinash Golande et al(2019)	Decision tree KNN K-mean clustering	86.60%
10.	Senthil Kumar Mohan et al(2019)	HRFLM- Hybrid Random Forest with a Liner Model	88.4%

3 Conclusion and Future Work

Different types of machine learning algorithms for prediction of heart disease have been summarized. The features have been analyzed and the various machine learning algorithms were studied and the best algorithm is chosen. Different result is given for every algorithm in different situations. Further it is analyzed, only marginal accuracy is achieved for heart disease prediction model. In order to increase the accuracy of predicting the heart disease earlier, the complex models must be used. In future, we will propose a methodology for early hear disease prediction to increase the accuracy to a higher rate and to reduce the cost and complexity to be minimum.

References

1. Shaikh Abdul Hannan, V. D. Bhagile, R. R. Manza and R. J. Ramteke. Diagnosis and Medical Prescription of Heart Disease Using Support Vector Machine and Feedforward Backpropagation Technique. International Journal on Computer Science and Engineering, 2018.
2. Poornima Singh, Sanjay Singh, Gayatri S Pandi-Jain,. Effective heart disease prediction system using data mining techniques. International Journal of Nano Medicine, 2018.
3. Mr.J.Santhana Krishnan, Dr.S.Geetha. Prediction of Heart Disease Using Machine Learning Algorithms, 2019. 1st International Conference on Innovations in Information and Communication Technology (ICIICT). DOI:10.1109/ICIICT1.2019.8741465.
4. Muhammad Waqar, Hassan Dawood, Hussain Dawood, Nadeem Majeed,Ameen Banjar, and Riad Alharbey. An Efficient SMOTE-Based Deep Learning Model for Heart Attack Prediction. Scientific Programming Volume 2021, January 2021.
5. Prashasti Kanikar and Disha Rajeshkumar Shah. Prediction of cardiovascular diseases using SVM and Bayesian Classification. International Journal of Computer Applications (0975 – 8887) Volume 156 – No 2, December 2016.
6. Avinash Golande, Pavan Kumar T. Heart Disease Prediction Using Effective Machine Learning Techniques. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019.
7. Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava. Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques. Digital Object Identifier 10.1109/ACCESS.2019.2923707. IEEE Access, VOLUME 7, 2019. S.P. Bingulac. On the Compatibility of Adaptive Controllers. Proc. Fourth Ann. Allerton Conf. Circuits and Systems Theory. pp. 8-16, 1994. (Conference proceedings).

8. Abdullah. A.S and Rajalaxmi.R.R, A data mining model for predicting the coronary heart disease using random forest classifier, "in Proceedings International Conference in Recent Trends Computer Methods, Communication. Controls, Apr. 2012, pp. 22–25.
9. A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, „Using PSO algorithm for producing best rules in diagnosis of heart disease," in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017, pp. 306– 311.
10. H. A. Esfahani and M. Ghazanfari, „Cardiovascular disease detection using a new ensemble classifier," in Proc. IEEE 4th Int. Conf. Knowl.- Based Eng. Innov. (KBEI), Dec. 2017, pp. 1011–1014