Journal of Clinical and Experimental Cardiology

OPEN ACCESS Freely available online

Research Article

Noninvasive Method for Heart Disease Prediction Using Machine Learning Algorithms for Photoplethysmograph Signals

Tripti R Kulkarni[1*], Dushyanth ND[2]

[1]Department of Electronics & Communication, Dayananda Sagar Academy of Technology and Management, Bengaluru, India;[2]Department of Electronics & Communication, K S School of Engineering and Management, Bangalore, India

ABSTRACT

Heart is a very crucial body organ that maintains and conjugates the blood in our body. In a report from WHO, one of the most significant causes of mortality today is Heart disease. Symptoms of a Heart disease include abnormal heartbeat, shortness of breath, pain in the chest, back or neck, fatigue and anxiety. If diagnosed well in advance, a lot of lives can be saved. In the proposed model we try to predict Heart Diseases in a patient using Machine Learning (ML). Various ML algorithms such as support vector classifiers, KNN, decision tree are being employed. Different ML Algorithms are used and accuracy is compared among them.

Keywords: Machine learning; Coronary artery disease; Decision tree; KNN; Support vector machine; PPG

INTRODUCTION

One of the foremost vital organs of the humanoid body is the heart. A very common cardiac illness in the world is Heart Attack. In an appraise by the World Health Organization, that over 17.9 million deaths occur each year around the world due to cardio vascular diseases, and of these passing, 80% are due to coronary artery illness and cerebral stroke [1]. In 2015, approximately 17.7 million people died due to Cardio-vascular diseases which are 31% of the total global deaths [2]. Change in the way of life, work stress and wrong food propensities add to several heart related ailments. Heart related sickness increases the cost of medical services and furthermore diminishes the effectiveness of a person. It is difficult to detect heart related ailments due to a few contributory risk factors, for example, diabetes, hypertension, elevated cholesterol, strange heartbeat rate and numerous other factors. Heart related ailments have made a lot of genuine worries among researchers; one of the significant difficulties in the heart diseases is its correct detection and discovering presence of it inside a human body. Early strategies have not been much effective in finding it even clinical researchers are less productive enough in predicating the heart related ailments.

Therefore, sensible and precise recognition of heart related ailments is significant. Clinical associations, all around the globe, gather information on different wellbeing related issues. This information can be mistreated using different machine learning techniques to gain helpful understandings. But the information gathered is exceptionally gigantic and many times, this information can be up roarious. These datasets, which are excessively wrecking for human personalities to grasp, can be effectively investigated using different machine learning strategies. Subsequently, these algorithms have gotten valuable, as of recent times, to anticipate the presence or absence of heart related diseases accurately.

This paper aims to analyze huge complex medical data to predict heart related ailments in a person. To get an efficient, accurate and early medical diagnosis of heart disease in taking preventive measures to prevent death. This paper does a deep analysis in the prediction of heart diseases with the help of Machine Learning algorithms. Machine Learning (ML) which is a sub field of data mining that deals with large scale well formatted dataset productively. Various Machine Learning algorithms such as K Nearest Neighbor, Support Vector Machine, Decision Tree, are compared to find the most accurate model [3]. These features are measured from the Photoplethysmograph (PPG) signals taken from PPG-BP data set [4]. These features are measured from the PPG signals taken from PPG-BP data set [4]. PPG signals collected from 219 subjects with different kinds of Cardio vascular related diseases and an age ranging from 20 to 89 years. This is an open clinical database, useful for the non-invasive detection of cardiovascular related diseases also heart disease data set from the UCI repository is being taken. In this research a discussion and examination of the existing classification techniques is made.

Correspondence to: Tripti R Kulkarni, Department of Electronics & Communication, Dayananda Sagar Academy of Technology and Management, Bengaluru, India, Tel: 8050048931; E-mail: kulkarnitripti@gmail.com

Received date: October 27, 2021; **Accepted date:** November 10, 2021; **Published date:** November 17, 2021

Citation: Kulkarni TR, Dushyanth ND (2021) Noninvasive Method for Heart Disease Prediction Using Machine Learning Algorithms for Photoplethysmograph Signals. J Clin Exp Cardiolog. 12:702.

LITERATURE SURVEY

Different researchers have contributed to the development of Machine Learning. Predictions of heart disease based ML algorithms is always a curious case for researchers. Lately there's a wave of papers and study fabrics in this area. The Medical Industry generates a large amount of data that has not been used effectively. Our aim in this proposed model is to bring out all state of art work by different authors and researchers.

Mohan et al. have used the UCI repository from the Clevel and Heart Disease database to collect the data so that Supervised Machine Learning techniques such as support vector classifiers, random forest, kNN, naïve bayes, decision tree and logistic regression can be employed. The missing values in the datasets will be handled and visualized. Then the accuracy obtained from various ML Algorithms is compared [5].

Kamboj [6] has used an approach known as the Exploratory Data Approach (EDA) to detect the main highlights in the datasets. It is done using visual methods. In the ML approach, EDA feature selection and engineering are usually done together and are considered very important.

Himanshu and Rizvi both are worked and [7] have proposed that data preprocessing is needed so that the crucial features can be obtained from the datasets. This helps to reduce data redundancy, inconsistency and the time complexity. An Electroencephalography (EEG) device is used to collect the data to train the model for pattern classification by the authors and have obtained an efficiency of up to 97% while using the SVM technique.

Nandhini et al. [8] had worked together and a model is proposed in which the system notifies in case of an emergency situation. Both diagnosis and monitoring of heart diseases takes place. The patients made to wear a smart watch fitted with a pulse rate sensor and Bluetooth which sends the data to the mobile application. Here, the data from the pulse rate sensor can also be pushed to the Cloud for further analysis of Heart Rate Variability (HRV). The chances of occurrence of the disease can be predicted by monitoring the increase or decrease in the HRV. Based on this, the user will receive a notification on the smart phone. For a new user, the presence or absence of the heart disease can be determined using the datasets and specified ML Algorithms.

Dashtipour, et al. [9] have jointly used neural networks and fuzzy rules. After training the neural network, fuzzy rules are applied which handle the truth values for being either completely true or completely false. This classified the diseases in a category of low, medium or critical but did not talk about the accuracy.

Nagamani, et al. has explained the outcomes by cleaning the datasets and transformation of data is done using preprocessing. A heart-disease-diagnosis attribute named as "num" is used. The presence of the disease is indicated using terms like low, medium, high and very high can be associated with num. The value of num would be zero in case of its absence.

Golande, Pavan [10] both are worked and the pre-processing is done to remove NaN (missing values marker) values to convert them into numerical values. This is followed by splitting the whole database in which 80% data is used for training and 20% is used for testing. The data is then trained using Supervised Machine Learning techniques.

Deep learning algorithms for Arrhythmia detection using PPG signal have not been widely applied. However, we found few studies for Arrhythmiaprediction, but when compared to the existing method the performance of the method presented in this paper is better, in terms of Accuracy, Sensitivity, Precision, FPR (False Positive Rate), FNR (False Negative Rate), Specificity, F1score [11-17].

MATERIALS AND METHODS

The PPG signals are taken from PPG-BP data set [11]. These features are measured from the PPG signals taken from PPG-BP data set [11] PPG signals collected from 219 subjects with different kinds of Cardio vascular related diseases and an age ranging from 20 to 89 years. This is an open clinical database, useful for the non-invasive detection of cardiovascular related diseases. The dataset contains NaN (Not a Number) values. The NaN values cannot be processed by the programming hence these values need to be converted into numerical values. In this approach mean of the column is calculated and NaN values are replaced by the mean.

Some features have a greater influence over the other. Hence normalizing the dataset and placing the values in a proper range using Min Max scalar/Standard Scalar. The dataset is split into training and testing dataset. 67% data is taken for training while remaining 33% data is used for testing. Below Figure 1 is the flow chart of the complete process.

Figure 2 represents that some features even have negative correlation with the target value, but few features have lower positive correlation.

The training data is trained by using four different machine learning algorithms i.e. Decision Tree, KNN, SVM. Each algorithm is explained in detail.

I) K-Nearest Neighbor (K-NN): The K-nearest neighbor's algorithm is a supervised classification algorithm method. It groups objects on the basis of nearest neighbors. The computation of distance of a feature from its neighbors is measured using Euclidean distance. A group of named points is being used to mark another point.

The data are clustered based on similarities between them, and is possible to fill the missing values of data using K-NN. Once the missing values are filled, various prediction techniques are applied to the data set as shown in figure 3. It is possible to gain better accuracy by utilizing various combinations of these algorithms. The KNN algorithm is easy to complete without creation of any model or making other assumptions. This algorithm is versatile and can be effectively used for classification, regression, and search. Even though K-NN is one of the simplest algorithms, noisy and irrelevant features affect its accuracy. In a study by Pouriyehetal, 83.16% accuracy was achieved with esteem K =9.

II) Decision tree: Decision tree is a classification algorithm that deals with both categorical as well as numerical data. Decision trees are used for creating tree-like structures. Decision tree is easy to implement and is greatly used to handle clinical dataset. The decision tree model examines based on tree nodes as it can be seen in figure 4.

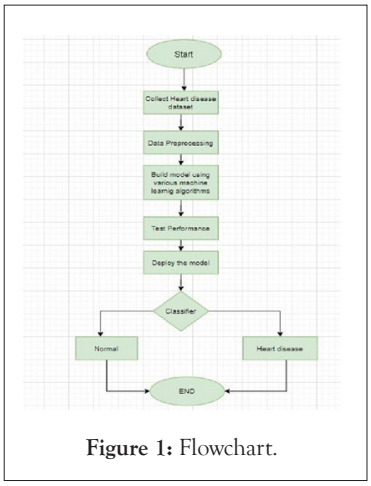

Figure 1: Flowchart.

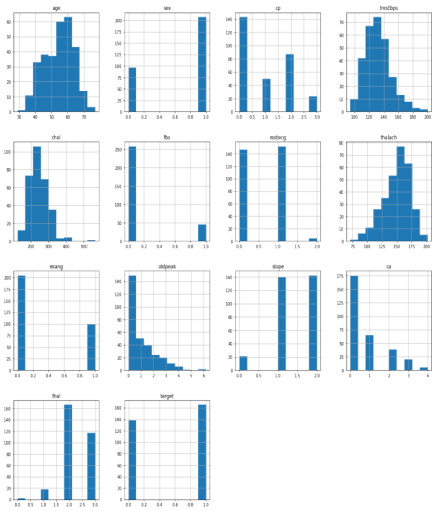

Figure 2: Histogram representation of various datasets.

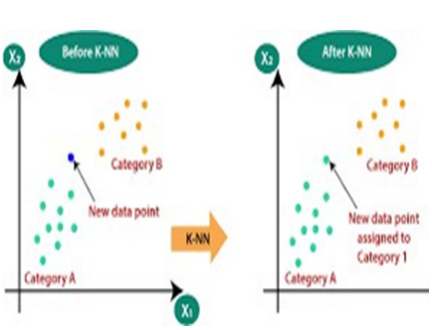

Figure 3: K-NN Algorithm.

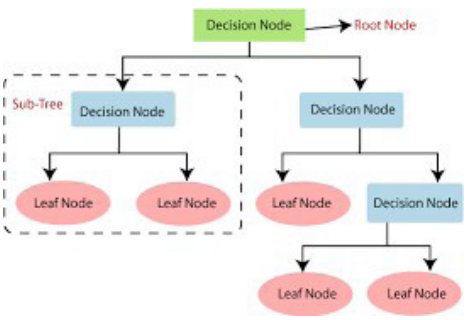

Figure 4: Decision tree.

This algorithm divides the data into at least two closely resembling sets dependent on the main predictors. The entropy of each feature is determined and then the data are divided, with predictors having maximum information gain or minimum entropy.

The results obtained are easier to peruse and decipher. This algorithm has higher precision in contrast to other algorithms as it analyzes the dataset in the tree-like graph. However, the data may be over classified and only one feature may be tested at a time for decision-making.

III) Support vector machine: A Support Vector Mechanism (SVM) is a discriminative classifier formally characterized by an isolating hyper plane as shown in Figure 5. Basically, we need to put the algorithm output in an optimal hyper plane which classifies new models. In 2-D space, this hyper plane is a line isolating a plane into two sections where in each class lay on one or the other side. The points whose positions are in the separating hyper plane are called Support Vectors. There are n number of ways to draw support vector and hyper plane.

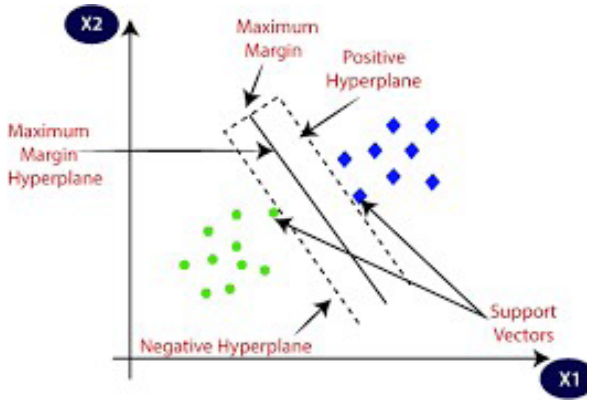

Figure 5: Support vector machine.

The distance between the Canonical and Separating hyper plane is called Margin. Sequential minimal optimization has been implemented which is a variant of SVM. The minimal sequential optimization breaks the issues in sub-problems and then solves it analytically.

RESULTS AND DISCUSSION

This research was done using technique of K-NN, Support Vector Machine (SVM), and Decision Tree using Kaggle open dataset. Dataset was splitted into 67% training and 33% testing the model. Python 3 using the platform Anaconda Navigator was used to note the accuracy. The performance of the various parameters used for classification purpose is shown in the tables below:

Total No. of Samples=303, Healthy Samples=138, Unhealthy Samples=16

The above Table 1 shows the count of people in different age groups who are normal and those who are suffering from heart disease.

Table 1: Age.

Age group	Normal	People suffering from heart disease
20-30	0	1
30-40	4	11
40-50	23	49
50-60	60	64
60-70	50	31
70-80	4	6

The above Table 2 shows the count of male and female without heart disease and with heart disease.

Table 2: Gender.

Gender	Without disease	With heart disease
Male	114	93
Female	24	72

Table 3 shows how different parameters are individually responsible for the chance of heart attack. For example there are 164 people who are having no chest pain and are normal but 139 people having chest pain are more proned to heart attack.

Table 3: Performance of different parameters.

SI No.	Parameter	Correctly classified	Wrongly classified	Accuracy%
1.	Chest pain	164	139	54%
2.	Resting blood pressure	147	152	49%
3.	Cholesterol	191	138	63%
4.	Fasting blood sugar	258	45	86%
5.	Resting ECG	182	121	60%
6.	Max. heart rate	44	259	14%
7.	Exercise induced angina	204	99	67%

Table 4 shows a comparison between the three algorithms.

Table 4: Performance of different parameters.

Algorithm	Accuracy
K-NN	83%
SVM	80%
Decision tree	78%

CONCLUSION

This project involves prediction of the heart disease dataset with proper data processing and implementation of machine learning algorithms. Preprocessing of a dataset for the removal of duplicate records, normalizing the values is used to represent information in the database. Among all the machine learning algorithms the one used in this project are K- NN, SVM, and Decision Tree algorithm. The features selection algorithms are used for features selection to increase the classification accuracy and reduce the execution time of classification system. The performance measuring metrics are used for assessment of the performances of the classifiers.

The performances of the classifiers have been checked on the selected features as selected by features selection algorithms. The suggested diagnosis system is aimed to achieve good accuracy as compared to previously proposed methods. Additionally, the proposed system can easily be implemented in healthcare for the identification of heart disease.

REFERENCES

1. Shah D, Patel S, Bharti SK. Heart disease prediction using machine learning techniques SN computer science. Springer. 2020; 1:345.
2. Malav A, Kadam K. A hybrid approach for heart disease prediction using artificial neural network and K-mean. International Journal of Pure and Applied Mathematics. 2018; 118:103-109.
3. Anbuselvan P. Heart disease prediction using machine learning techniques. IJERT. 2020; 9:1-4.
4. Liang Y, Liu G, Chen Z, Elgendi M. PPG-BP Database. 2018.
5. Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE Journal. 2019; 7: 1-13.
6. Kamboj M. Heart disease prediction with machine learning approaches. IJSR. 2020; 9:1454 – 1458.
7. Sharma H, Rizvi MA. Prediction of heart disease using machine learning algorithms: a survey. International Journal on recent and Innovation Trends in Computing and Communication. 2017.
8. Nandhini S, Debnath M, Sharma A, Pushkar. Heart disease prediction using machine learning. IJRERD. 2018; 3:1-8.
9. Dashtipour K, Ieracitano C, Morabito FC, Raza A. Hussain A. An ensemble based classification approach for persian sentiment analysis. In Progresses in Artificial Intelligence and Neural Systems. 2021; pp.207-215.
10. Golande A, Pavan T. Heart disease prediction using effective machine learning techniques. IJRTE. 2019; 8:1-7.
11. Nagamani T, Logeswari S, Gomathy B. Heart disease prediction using data mining with mapreduce algorithm. IJITEE. 2019; 8:1-4.
12. Gilon GJ, Subramanian CN, Bersini H. Forecasting episodes of atrial fibrillation using a new machine learning algorithm Arch. Cardiovasc. 2020; 12:103.
13. Aschbacher K, Yilmaz D, Kerem Y, Crawford S, Benaron D, Liu J, et al. Atrial fibrillation detection from raw photoplethysmography waveforms: A deep learning application. Heart Rhythm. 2020; 1:3-9.
14. Xiong Z, Liu T, Tse G, Gong M, Gladding PA, Smaill BH, et al. A machine learning aided systematic review and meta-analysis of the relative risk of atrial fibrillation in patients with diabetes mellitus. Front Physiol. 2018; 9: 835.
15. Kulkarni TR, Dushyanth ND, Early and non-invasive screening of common cardio vascular related diseases such as diabetes and cerebral infarction using photoplethysmograph signals. Elsevier. 2021;3: 1-13.
16. Kulkarni TR, Dushyanth ND. Performance evaluation of deep learning models in detection of different types of arrhythmia using photo plethysmography signals. Springer. 2021.
17. Sridhar A, Kapardhi AS. Predicting Heart disease using machine learning algorithm with SDNKL. IRJET. 2020; 7:1-7.