IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Detailed Survey Of Machine Learning Algorithms- ECG For Heart Related Diseases

Dr. H S Prasantha
Professor
Department of Computer Science and Engineering
K.S. Institute of Technology, Bangalore, India

***Abstract:*** Machine learning (ML) is a subset of artificial intelligence (AI) where computer is trained to make decisions like a human based on different characteristics from a set of data. Machine Learning is a subset of artificial intelligence (AI) which allows the software applications to predict the outcomes by considering the historical data as input. The paper presented discusses the detailed understanding and the contribution of different researchers in the area of Biomedical Signal processing.

***Keywords***- ECG, machine learning, Heart diseases

## 1. Introduction

Biomedical Signal processing is the domain which is used to process the signal from biomedical sensors for monitoring and identification of health related issues. Electrocardiography is the process of producing the recording of variations of heart rate in the form of electrical signal. Changes in the normal ECG pattern occur in numerous cardiac abnormalities, including cardiac rhythm disturbances. They are time varying signals plotted as P, Q, R, S and T. P wave shows depolarization of right and left atrium. Q,R, and S waves occurs successively as QRS complex which represents the electrical flow in the ventricles and implies right and left ventricular depolarization. T wave indicates ventricular repolarisation.

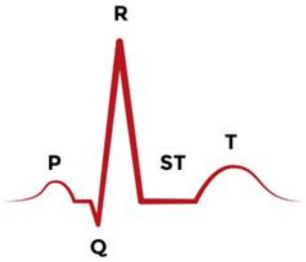

Figure 1: basic ECG signals with P,Q,R,S and T waveforms

Machine learning can help us realize improved healthcare by unlocking the potential of large biomedical and patient dataset. Machine learning techniques have been used to predict cardiovascular (CV) events, including the risk of incident HF.The objective of this case study is to check whether the patient is likely to be diagnosed with any cardiovascular heart diseases. Attributes of heart disease are useful in developing ECG applications that would automatically detect heart disease. It is also useful in selecting suitable datasets to use in training ML models for prediction or classification of heart disease. Selecting the attributes play a very important role in training our model. Features or attributes that are distinct and discriminating are selected for training our model. Some of the common attributes used for diagnosing heart diseases are age, sex, resting blood pressure, chest pain type, resting ECG, cholesterol levels, fasting blood sugar. These are few attributes that could differentiate a person with normal healthy heart and that of a person who could have a heart disease. Dataset collected are divided into three main parts : training data ,validating data and testing data. Usually, seventy percent of the data is used for training the model, twenty percent for validating the model and remaining ten percent for testing the model. If the dataset is not utilized in a proper manner, our model could over fit and affect the performance and accuracy of the model.

## 2. LITERATURE SURVEY

The efforts are made in the paper to identify the contributions made by different authors related to identification and prediction of Heart related issues. The detailed discussion is made with respect to the approaches/methodology adapted and the outcome of the work.

The paper (1) titled detection of cardiovascular diseases in ECG Images Using Machine Learning and Deep Learning Methods uses Naive Bayes (NB) approach. The author proposed a lightweight CNN-based model to classify the major cardiac abnormalities. The proposed approach by the author using Naïve Bayes approach resulted in a accuracy rate of 99.79%.

The paper (2) titled Classification of ECG signals using machine learning techniques: A survey uses Discrete Wavelet Transform, Continuous Wavelet Transform, Discrete Cosine Transform and Tompkins's algorithm. From the paper we can conclude that ECG data are classified into ECG beat classification and ECG signal classification. For ECG Beat classification, many researchers have used the MIT-BIH arrhythmia database and neural networks as a classifier. Moreover, it is observed from survey that MLPNN gives good accuracy for ECG beat classification.

The paper (3) titled Analysis and classification of heart diseases using heartbeat features and Machine learning algorithms uses the algorithms such as Decision tree algorithm, Gradient Boosted Tree (GBD) algorithm and Random forest algorithm. The outcome of the work showed an accuracy of 96.75% using GBD algorithm, 97.98% using random forest for binary classification and For multicast accuracy of 98.03%using random forest.

The paper (4) titled A Study on ECG Signal For early detection of heart diseases using machine learning technique uses three different training algorithms such as "trainingdx", "trainrp" and "trainlm" algorithms using MATLAB Based tool. In this paper, author discussed about various machine learning techniques used to detect arrhythmia type. It was observed that SVM models outperformed Neural Networks when number of arrhythmia types to be detected is more. But SVM models are influenced by feature selection, whereas Deep learning models such as CNN can extract their own features. Recently, Convolution Neural Network and Long Short-Term Memory models from deep learning are becoming increasingly popular and showing promising results in arrhythmia detection from ECG

The paper (5) titled Computational Diagnostic Techniques for Electrocardiogram Signal Analysis uses Machine learning algorithms including K nearest neighbour (KNN), decision tree, random forest, logistic regression, support vector machine (SVM), naive Bayes, K mean algorithm, ADA boost algorithm, neural network, Markov, and so on. The contribution is providing summary of the latest computational diagnostic techniques based on ECG signals for estimating CVD conditions. The classic machine learning techniques play important roles in efficient and reliable monitoring of the ECG activity in hospital settings or at home by analyzing ECG recordings. The procedure of ECG signals analysis is discussed in several subsections, including data pre-processing, feature engineering, classification, and application.

The paper (6) titled Machine-Learning based Heart Disease Diagnosis: A Schematic Lecture Review uses SVM, CNN, DNN, ANN, MLP etc. The contribution of the work is the study of Systematic Literature Review (SLR) approach to uncover the challenges associated with imbalance data in heart diseases prediction. It deals with model's performance while disregarding issues like interpretability and explain ability.

The paper (7) titled Application of Machine Learning for the Detection of Heart Disease uses Fuzzy KNN and BP-Neural Network. The author Work conducted a study of different algorithms. Compared to the success of K Means Clustering, KNN, logistic regression, etc. BP-Neural Network and Fuzzy KNN appear to be better performance with rate of 98.00% and 94.19%.

The paper (8) titled early detection of heart diseases using a low-cost compact ECG sensor uses CNN, Random Forest, Gradient Boosting algorithm. The contribution is the Creation of a new data set for classifying healthy and diseased heart patients. The paper improves the classification accuracy by handling numerous data set imbalance techniques. 1D CNN with the support of the oversampling technique and with voting strategy gave the best result. i.e., 93%.

## 3. MACHINE LEARNING ALGORITHMS

DECISION TREE ALGORITHM: Decision Tree is a supervised algorithm that can be used for classification and regression problems. It is a tree-structured classifier which has internal nodes which represent features that are in the database, branches which represent the decision rules and each leaf node represent the final outcome. There are two nodes namely decision node which has multiple branches that helps us with decision making, leaf node which are the output of the decision made. In classification problem the final result is either of the one class in case of binary classification. Decision tree is drawn upside down with its root at the top. Each node in the tree specifies a test of some attribute of the instance, and each branch descending from that node corresponds to one of the possible values of the attribute. An instance is classified by starting at the root node of the tree, testing and then moving down the tree according to the value of the attribute given. To select which attribute acts as the root node we consider evaluating it statistically by using entropy, gain etc.

**RANDOM FOREST ALGORITHM:** Random Forest algorithm is a supervised machine learning algorithm that is mostly used in classification and regression problems. Random forests are a parallel combination of decision trees, which use bagging type of learning. It is based on ensemble learning, which is used to improve the performance and uses multiple classifiers to solve a complex problem. It uses several decision trees as subsets to increase the accuracy .the greater the decision trees used higher the accuracy and better the performance. This algorithm also makes sure that the model does not overfit and regulates it. Since it uses multiple different decision trees, few decisions tress can predict correctly while few cannot. Therefore, we make two assumptions: one, there should be some actual values in the feature variable of the data set so the classifier can predict correct results rather than choosing a random integer and guessing them. Two, the predictions from each tree must have very less correlations. It takes less time to train the model .The main two design steps involved are : first to create random forest by combining multiple decision trees and second is to make predictions for each tree created. Thus, random forest can be used to train our model for better accuracy.

**GRADIENT BOOST TREE ALGORITHM**: Gradient boosting tree algorithm is a supervised machine learning algorithm used for classification and regression problems. It works by building simpler models which can make predictions where each model tries to predict the error left over by the previous model. Ensemble learning is a model that makes predictions based on a number of different models. By combining these simpler models, a model which uses this type of learning are more flexible. The two most popular ensemble learning are bagging and boosting. Bagging is training models parallelly, each model learns from random subset of data. Boosting is training models sequentially; each model learns from the mistakes of the previous model. . Gradient boost algorithm is an application of boosting type of learning. Here we combine many weak learning models to make them as one strong learning model. The trees are connected serially, and each tree tries to predict and classify more accurately and minimize the error of the previous tree in the series. Due to sequential learning, this algorithm learns slowly but is highly accurate .The new learners' fits into the residuals of the other weak models, finally the model aggregates the result of each tree and thus stronger model is achieved.

**SUPPORT VECTOR MACHINE**: Support vector machine is one of the most popular supervised learning algorithms, which is used for classification as well as regression. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyper plane. SVM chooses the extreme points or vectors that help in creating the hyper plane. These extreme cases are called as support vectors and hence algorithm is termed as Support Vector Machine. There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyper plane of SVM.

**KNN:** K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique .KNN algorithm assumes the similarity between the new case/ data and available cases and put the new case into the category that is most similar to the available category. KNN algorithm stores all the available data and classifies the new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using KNN algorithm. It can be used for regression as well as for classification but mostly it is used for the classification problem.

**LOGISTIC REGRESSION:** Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.Logistic Regression is much like the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.

## 4. TOOLS

Python is a programming language that supports wide range of applications. One such application is building and training a machine learning (ML) model. Most of the AI, ML and deep learning models are built using python. Python has libraries that can support and train ML models. Python libraries can process data, analyse data, and can also help us to visualize them. Below are few libraries that can help us to train ML models.

**Tensor flow**: Tensor flow is one of the most popular open-source libraries used to train and build machine learning and deep learning models. It was developed by Google Brain team. It offers powerful libraries, tools, resources. It helps us to build and train AI/ML models. It provides with various abstracts that lets us use the resources according to our requirement.

**NumPy :**NumPy is well-known general-purpose array processing package. It can process large multi-dimensional arrays and matrices. It is also useful for handling linear algebra, Fourier transforms etc. Other libraries like TensorFlow etc. use NumPy in the backend. It can easily integrate most databases.

**SciPy:** this library offers modules for linear algebra, image optimization integration, interpolation, FFT, signal and image processing, differential equation solutions. SciPy uses NumPy internally for data and array manipulations. It can handle classification, regression, clustering pre-processing, dimensionality reduction etc.

**Theano:**Theano is a python machine learning algorithm that can optimize compiler for evaluating and manipulating mathematical expressions and matrix calculations. It can automatically avoid bugs and errors when dealing with logarithmic and exponential functions.

**Keras**: It is an open source used for neural networks and ML. Keras works with networks to build layers, activation functions and optimizers.it runs alongside with TensorFlow and Theano.

**Pandas**:Pandas are one of the most popular python libraries that is mostly used to analyze data. It is fast, flexible, and expressive. It can work on both labelled and unlabeled data. It can process matrix data with homogeneous or heterogeneous type of data in rows and columns.

**Matplotlib:**Matplotlib is a data visualization library that is used for 2D plotting to produce quality image plots and in different formats.

**PyTorch**: It has wide range of tools that can support computer vision, machine learning and NLP. It has smooth integration with python data science stack .It can perform computation on tensors.

ECG hardware has evolved from a unit that is big in size, wired to smaller wireless and wearable, which allows real time continuous monitoring of patients. Leads are the electric diodes that detects the change in electric flow. Here an electrode comprises of electric pad that connects the skin and allows the recording of electric current. Heart rate monitor boards assists in continuous measurement and displaying patients heart rate. Several processing boards are utilized to process the ECG signals. Most commonly used processing boards are Raspberry's pi and Arduino. The wireless communication capability of an ECG sensor is significant and enables the ECG sensors to transmit ECG signal recordings to nearby devices. The most commonly used communication standards used in ECG sensors are ZigBee, Bluetooth, and Medical Implant Communication service. Diagnoses, interpretation process, AI, ML, cloud computing and smart phone-based applications are used for practical real time applications

## 5. APPLICATIONS

Heart disease prediction is one of the most complicated tasks on medical field. As heart disease prediction is complex, there is a need to automate the prediction process to avoid risks associated with it and alert the patient in advance .Therefore, predicting heart disease could reduce the mortality rate thereby using this automated system at hospitals could help patients for early detection and cure.

## 6. REFERENCES

1. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0244-x: Analysis and classification of heart diseases using heartbeat features and machine learning algorithms.
2. https://ieeexplore.ieee.org/abstract/document/9074954 : applications of machine learning for the detection of heart disease
3. https://ieeexplore.ieee.org/abstract/document/7164783: classification of ECG signals using ML techniques.
4. https://www.researchgate.net/publication/358897643_Machine_Learning_Technology-Based_Heart_Disease_Detection_Models:Machine Learning technology – heart disease detection models.
5. https://www.researchgate.net/publication/356084524: A study on ECG signals for early detection of heart diseases using ML techniques.
6. https://www.sciencedirect.com/science/article/pii/S0933365722000549: machine learning – based heart disease diagnosis
7. https://ieeexplore.ieee.org/document/9735300: detection of cardiovascular diseases in ECG images using ML and deep learning methods.
8. https://link.springer.com/article/10.1007/s11042-021-11083-9:early detection of heart disease using a low-cost compact ECG sensor
9. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7664289/: computational diagnostic techniques for ECG signal analysis
10. Dr. H S Prasantha, "NOVEL APPROACH FOR IMAGE COMPRESSION USING MODIFIED SVD", International Journal of Creative Research Thoughts (IJCRT), Volume 8, Issue 8, Page 2234-2243, Aug 2020
11. Dr. H S Prasantha, "IMPLEMENTATION OF IMAGE COMPRESSION USING FAST COMPUTATION OF SVD ON DM642", International Journal of Creative Research Thoughts (IJCRT), Volume 8, Issue 8, Page 2364-2368, Aug 2020
12. Prasantha, H, H Shashidhara, K N B Murthy, and M Venkatesh. "Performance Evaluation of H.264 Decoder on Different Processors." International Journal on Computer Science & Engineering. 1.5 (2010): 1768. Web. 7 Apr. 2013.
13. H. S. Prasantha, H. L. Shashidhara, and K. N. Balasubramanya Murthy. Image compression using SVD. In Proceedings of the International Conference on Computational Intelligence and Multimedia Applications, pages 143–145. IEEE Computer Society, 2007.
14. Gunasheela K S, H S Prasantha, "Compressive sensing for image compression: survey of algorithms", Proceedings of Emerging Research in Computing, Information, Communication and Applications, ERCICA, Springer publication, Bengaluru, 2018
15. K N Shruthi, B M Shashank, Y. SaiKrishna Saketh, H.S Prasantha and S. Sandya, "Comparison Analysis Of A Biomedical Image For Compression Using Various Transform Coding Techniques", IEEE, pp. 297-303, 2016