IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10698

PREDICTION OF CARDIAC DISEASES BASED ON ECG ANALYSIS USING MACHINE LEARNING

Prabhavathi K[1], Kamal Kumar K[2], Priyanka M K[3], Manushree[4], Madhusmitha K G[5]

Assistant Professor, Electronics and Communication, BGS Institute of Technology, Mandya, India[1]

Student, Electronics and Communication, BGS Institute of Technology, Mandya, India[2-5]

Abstract: Heart arrhythmia may be a heart state during which the heartbeat is irregular. The heartbeat possibly is excessively quick, excessively slow, or unstable. Electrocardiography (ECG) is utilized for the detection of heart arrhythmia. Since ECG signals reflect the physiological states of the heart, specialists use ECG signs to analyze heart arrhythmia. Thus, the advancement of programmed procedures for recognizing strange states of ECG signals from the everyday recorded ECG information is of crucial significance. Additionally, ideal medical aid measures can be successfully applied if such unusual heart conditions can be distinguished naturally utilizing the wellbeing observing gear which will inside utilize the Machine Learning algorithms.

Keywords: Electrocardiography, ECG, Machine Learning, cardiac arrhythmia

I.INTRODUCTION

Heart arrhythmia is a typical manifestation of heart infections. A few sorts of heart arrhythmia like atrial fibrillation, ventricular break, and ventricular fibrillation may cause strokes and heart failure. The rhythm of a heartbeat is constrained by electrical motivation created in the sinoatrial node. An arrhythmia beat happens when there are a few problems in the typical sinus rhythm. Various arrhythmias can cause distinctive ECG designs. The arrhythmias, for example, ventricular just as atrial fibrillations and vacillates are hazardous and may likewise prompt stroke or unexpected cardiovascular demise. There are more prospects of arrhythmic pulsates for the patients who have recently experienced a coronary failure and further the high dangers of hazardous heart rhythms. Coronary illness stays the main source of death across the world in both metropolitan and provincial regions. The most well-known kind of coronary illness is a coronary illness which will bring about killing 380,000 individuals yearly. This is a supervised learning issue. These machine learning methods can be conveyed in emergency clinics where a huge dataset is accessible and it can help the specialists in settling on more exact choices and it even assists with chopping down the number of causalities because of heart sicknesses later on. After suitable element choice, we intend to tackle this issue by utilizing Machine Learning Algorithms Namely K Nearest Neighbor, Logistic Regression, Naïve Bayes, and SVM. These AI methods can be conveyed in medical clinics where huge datasets are accessible and can help the specialists in settling on a more exact choice. This is a characterization method dependent on the Bayes Theorem with an assumption of freedom among indicators. We have executed our own Naive Bayes binomial and multinomial classifiers in Python. We utilize the Naive Bayesian condition to figure the back likelihood for each class. The class with the most elevated back likelihood is the result of the forecast. In the principal case, the preparation testing information was parted 80% - 20% and in the subsequent case, the preparation testing information was parted 80% - 20 %. In SVM, a hyperplane is chosen to best separate the focuses in the info variable space by their class, either class 0 or class 1. In two measurements you can imagine this as a line. You can make characterizations utilizing this line. By connecting input esteems into the line condition, we figure whether another point is above or beneath the line. We have attempted both the polynomial and the straight bits for the SVM and discovered that the linear kernel outperformed the polynomial kernel. First and foremost, we eliminated a portion of the all-out highlights that were 95% of the time showing either all 0's or all 1's. On the off chance that any preparation occasion has a missing incentive for a given property, we set it as the mean of the worth give or take the standard deviation for that quality identified with the class it has a place with. Assuming for a given trait greater part of qualities are missing, we will dispose of that property and eliminate it from our preparation set. Learning Algorithms specifically K Nearest Neighbor, Logistic Regression, Naïve Bayes, and SVM. The second evenhanded of this undertaking is to foster a technique to powerfully arrange an ECG follow into one of 13 wide arrhythmia classes. We report our presentation for every one of the five techniques utilizing two distinct philosophies. We show results for every calculation, just as fluctuate different boundaries for better outcomes.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10698

II.ECG DATABASE

ECG is the strategy for estimating the electrical possibilities of the heart to analyze heart related issues. It is non-obtrusive, simple to secure and gives a helpful intermediary to infection conclusion. It has for the most part been utilized for arrhythmia recognition using the huge number of freely accessible ECG information bases.In current examination, freely accessible Physio Net, MIT–BIH arrhythmia information base tested at 360 Hz is utilized. Further, pulses from the whole dataset are arranged into five arrhythmia classes as suggested by ANSI/AAMI EC57:1998 standard. The MIT-BIH information base contains 48 records. Each record has length of 30 minutes with testing recurrence of360 Hz. These records are chosen from 24 hours accounts of 47 unique people. Our examination is centered around the grouping of four heartbeat classes in the MIT-BIH arrhythmia data set: Normal beat (N), Left pack branch block (LBBB), Right group branch block (RBBB), Premature ventricular withdrawal (PVC). Table 1 shows the appropriation of these heartbeat types among the different ECG accounts present in the data set.

Heartbeat	Type ECG Recording Containing Respective Type
N	100,101,105,112,115,000,000
LBBB	109,111,207,214
RBBB	124,212,231,232
PVC	105,109,116,119,214,000,000

Table -1: Distribution of heartbeats

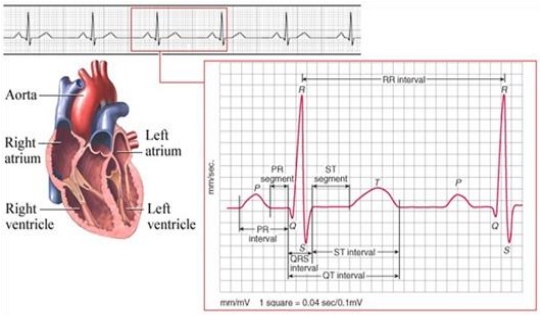


Figure 1. Components of ECG signal

The ordinary ECG signals are made out of P wave, QRS complex followed by T wave as displayed in Figure 1. Diagnosing of the heartbeat is relies upon researching the shape, the connection between these waves and the length of each wave. In any case, investigation of the heart state or ordinary ECG waves is certifiably not a simple assignment. Truth be told, the ECG signal is nonstationary and consequently, side effects of an infection, assuming any, may not happen routinely. In this manner, doctors need to record and screen the heartbeat for quite a while to arrange the mood into typical or strange sort. For ECG signal investigation, the size of the created information can be gigantic which requires a great deal of time and exertion, hence need for a programmed characterization framework.

III.MACHINE LEARNING

Being a field of science, Machine learning manages the manners by which machines will master utilizing their previous experience. According to viewpoints of numerous researchers, the expression "Machine learning" is conversely utilized alongside term "artificial intelligence", given that the chance of learning is the primary quality of an element called intelligent agent. The fundamental motivation behind machine learning is the development of PC program that can learn and adjust appropriately utilizing its previous experience. A more itemized and formal meaning of AI is given by Mitchel: A PC program is said to gain as a matter-of-fact E as for some class of undertakings T and execution measure P, if its presentation at assignments in T, improves with experience E, as estimated by P. With the coming of new methods and approaches of Machine Learning, we at present have a capacity to discover an answer for wellbeing related issues like heart arrhythmia. We can foster a framework utilizing Machine learning strategies which can distinguish if the patient has arrhythmia. Besides, recognizing the arrhythmia in a beginning phase, prompts treating the patients before it gets basic. AI can extricate concealed information from an enormous measure of heartbeat-related ECG information. Thus,

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10698

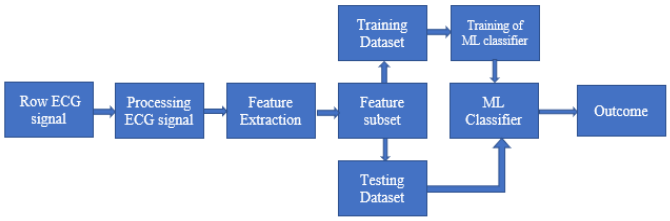
it's anything but a huge part in heart arrhythmia research, presently like never before. The point of this exploration is to foster a framework which can anticipate the sort of arrhythmia for a patient with a higher exactness.

IV.SUPERVISED LEARNING

This review is centered around the investigation of a framework dependent on characterization strategies that utilization managed learning. In administered learning; the framework must "learn" inductively utilizing a capacity called target work. This objective capacity is an outflow of a model which depicts information. The target work is utilized to anticipate the value of a variable, called the dependent variable or output variable, from a bunch of variables, called independent variables or input variables or characteristics or features. The arrangement of input of function as a contribution of the capacity, i.e., its area, are called instances. Each case is portrayed by a bunch of characteristics (attributes or features). A subset, all things considered, for which the yield variable worth is marked with the class it has a place, is called preparing information or models. To derive the best objective outcome, the learning framework, given a preparation set, should take a suitable speculation for the capacity and mean it by h. In directed learning, there are two sorts of learning errands: order and relapse. Order models anticipate unmistakable classes utilizing its prepared information, for example, e.g., blood gatherings, while relapse models foresee mathematical qualities. Probably the most well-known strategies are Support Vector Machines (SVM), Decision Trees (DT), Genetic Algorithms (GA), Artificial Neural Networks (ANN), and Instance-Based Learning (IBL, for example, k-Nearest Neighbors (k-NN).

V.METHODOLOGY

The framework design of the proposed framework is as displayed in Figure. Contribution to the framework will be a raw ECG signal. This raw sign contains noise. Pre-preparing of ECG signal eliminates this noise. Three distinctive denoising procedures which can be utilized are median filter, moving average filter, and notch filter. After these features are extricated from the channel ECG signal.



All out 9 highlights are extricated for each beat utilizing discrete wavelet change, specifically R point area, region under QRS complex, length of QR, RS, RR focuses, R top, R ordinary, region under autocorrelation, and SVD of ECG. Various methods i.e., FFT, CWT, and DWT, and so forth will be utilized for the extraction of various highlights from the denoised ECG signal. The subsequent element dataset will be then separated into a preparation dataset and a testing dataset. The preparation dataset will be feed to the diverse Machine Learning classifiers. In the proposed framework a SVM classifier and an ANN classifier will be utilized. Various theories and loads will be mulled over to expand the exactness of order. Eventually, the best blend of the pre-handling and arrangement strategies coming about because of leading this investigation will be utilized which can most precisely recognize the kind of heart arrhythmia. For execution assessment, we utilized three standard measurements in particular sensitivity, specificity, and accuracy. These measurements are utilized to evaluate the exhibition of the framework. The sensitivity is a proportion of the ability to test the positive examples.

Sensitivity,
(Sn) = (TP/TP + FN) * 100
Where TP addresses the genuine positive and FN addresses the bogus negative. The explicitness is a proportion of the ability to test the negative examples.
Specificity,
(Sp) = (TN/TN + FP) * 100
Where TN addresses the genuine negative and FP addresses the false positive. Accuracy is characterized as the capacity of the test to accurately recognize a classified type with and without positives. It reflects both sensitivity and specificity.
Accuracy,
(Ac) = (TP + TN)/ (TP + TN + FP + FN) * 100

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 10, Issue 6, June 2021
DOI 10.17148/IJARCCE.2021.10698

A. Modules used

- NumPy
- pandas
- matplotlib
- scikit-learn
- seaborn
- xgboost

B. Datasets

1.Age: Patients Age in years
2.Sex: Gender of patient (Male - 1, Female - 0)
3.Chest Pain Type: Type of chest torment experienced by understanding ordered into 1 normal, 2 ordinary angina, 3 non-anginal agony, 4 asymptomatic
4.Resting bps: Level of pulse at resting mode in mm/HG
5.Cholesterol: Serum cholesterol in mg/dl
6.Fasting glucose: Blood sugar levels on fasting > 120 mg/dl addresses as 1 if there should be an occurrence of valid and 0 as bogus
7.Resting ECG: Result of an electrocardiogram while very still are addressed in 3 particular qualities 0: Normal 1: Abnormality in ST-T wave 2: Left ventricular hypertrophy
8.Max pulse: Maximum pulse accomplished
9. Practice angina: Angina actuated by practice 0 portraying NO 1 portraying Yes
10.Old pinnacle: Exercise-actuated ST-despondency in examination with the condition of rest
11.ST incline: ST-portion estimated as far as the slant during top exercise 0: Normal 1: Upsloping 2: Flat 3: Down inclining
Target variable
12.Target: It is the objective variable which we need to anticipate 1 method the patient is experiencing heart hazard and 0 methods the patient is ordinary.

C. K-nearest neighbors (k-nn)

K Nearest Neighbor is a straightforward calculation that stores every one of the accessible cases and characterizes the new information or case dependent on a similitude measure. It is for the most part used to arrange an information point dependent on how its neighbors are characterized.

In the characterization setting, the K-closest neighbor calculation basically reduces to shaping a larger part vote between the K most comparative cases to a given "concealed" perception. The likeness is characterized by a distance metric between two information focuses. A well-known one is the Euclidean distance technique. KNN can be utilized for both characterization and relapse prescient issues. Nonetheless, it is all the more generally utilized in arrangement issues in the business.

- In k-NN characterization, the yield is class participation. An item is characterized by a majority vote of its neighbors, with the article being doled out to the class generally normal among its k closest neighbors (k is a positive whole number, commonly little). Assuming k = 1, the article is basically appointed to the class of that solitary closest neighbor.
- In k-NN relapse, the yield is the property estimation for the item. This worth is the normal of the upsides of k closest neighbors.
- k-NN is a sort of grouping where the capacity is just approximated locally and all calculation is conceded until work assessment. Since this calculation depends on distance for arrangement, if the highlights address distinctive actual units or come in endlessly various scales then, at that point normalizing the preparation information can further develop its precision drastically
- Both for classification and regression, a useful technique can be to assign weights to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. For example, a common weighting scheme consists of giving each neighbor a weight of 1/d, where d is the distance to the neighbor.

D. Random forest classifier

Random Forest Classifier is a troupe calculation. In the following couple of posts, we will investigate such calculations. Ensembled calculations are those which joins more than one calculation of the equivalent or diverse kind for arranging objects. Irregular backwoods is a regulated learning calculation. The "timberland" it's anything but, a group of choice trees, typically prepared with the "packing" strategy. The overall thought of the stowing strategy is that a mix of learning models builds the general outcome.

One major benefit of arbitrary woodland is that it very well may be utilized for both characterization and relapse issues, which structure most of current AI frameworks.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10698

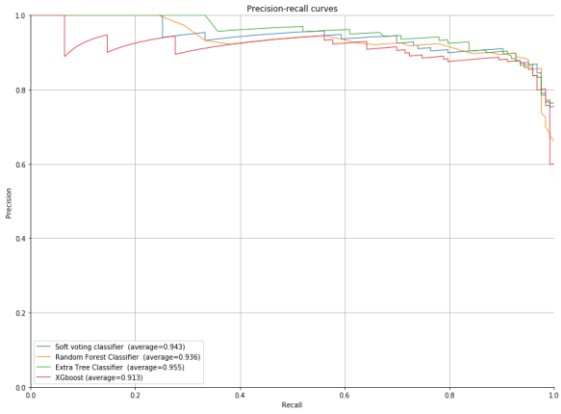


Figure 2: Precision recall curves

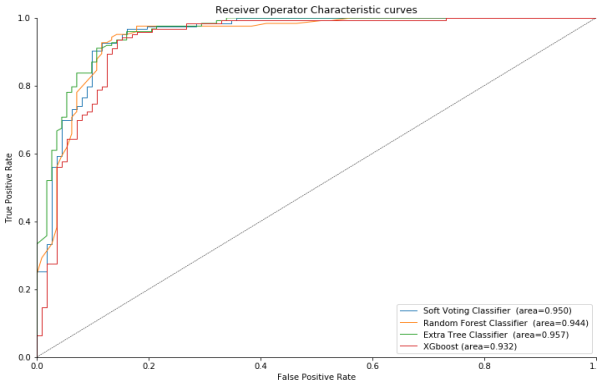


Figure 3: Receiver Operator characteristic curves

VI.CONCLUSIONS AND FUTURE WORKS

In the proposed system, random forest algorithm is used to obtain the high accuracy based on feature and pattern selection. From the above work, it is seen that the execution of ML-based procedures for coronary illness recognizable proof is working on the exactness and decreasing the expense factor. Nearly, everything recognizes the chance of cardiovascular breakdown without any significant clinical foundation hardware yet with canny ML methods. Utilizing arbitrary backwoods, we assembled a stage that can be utilized to characterize the heart patients with a accuracy of 90.21%. As an extension to the present work and a few kind of limitation to the work performed here, differing types of classifiers are often included within the analysis and more in-depth sensitivity analysis are often performed on these classifiers, also an extension are often made by applying same analysis to other bioinformatics diseases' datasets, and see the performance of those classifiers to classify and predict these diseases.

REFERENCES

[1] Padmavathi K, Ramakrishna K S, (2015), ―Classification of ECG signal during Atrial Fibrillation using Autoregressive modeling‖,Procedia Computer Science,Vol.46, IssueICICT 2014,pp.53-59, 2015.
[2] Sharma A, Bhardwaj K, (2015), ―Identification Of Normal And Abnormal ECG Using Neural Network‖,International Journal of Information Research and Review, Vol.2, Issue No.05, pp.695-700, May 2015.
[3] Subbiah S, Patro R K, Subbuthai P, (2015),‖Feature Extraction and Classification for ECG Signal Processing based on Artificial Neural Network and Machine Learning Approach‖,International Conference on Inter Disciplinary Research in Engineering and Technology [ICIDRET], Vol.1, Issue. ICIDRET007, pp.50-57, March 2015.
[4] Huang G, Huang G B, Song S, You K, (2015), ―Trends in extreme learning machines: A review‖,Neural Networks, Vol.61, pp.32–48,2015.
[5] Afkhami R G, Azarnia G, TinatiMd A, (2016), ‖Cardiac Arrhythmia Classification Using Statistical and Mixture Modeling Features of ECG Signals‖, Pattern Recognition Letters, Vol.70, pp.45-51,2016.
[6] Halil İbrahim BÜLBÜL and Neşe USTA, "CLASSIFICATION OF ECG ARRHYTHMIA WITH MACHINE LEARNING TECHNIQUES", 2017 16th IEEE International Conference on Machine Learning and Applications, DOI 10.1109/ICMLA.2017.0-104
[7] Juyoung Park, Seunghan Lee, and Kyungtae Kang, "Arrhythmia Detection using Amplitude Difference Features Based on Random Forest", 978-1-4244-9270- 1/15/$31.00 ©2015 IEEE

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10698

[8] Shalin Savalia and Vahid Emamian, "Cardiac Arrhythmia Classification by Multi-Layer Perceptron and Convolution Neural Networks", Bioengineering 2018, 5, 35; doi:10.3390/bioengineering5020035
[9] Sandipan Chakroborty and Meru A. Patil, "Real-time Arrhythmia Classification for Large Databases", 978-1- 4244-7929-0/14/$26.00 ©2014 IEEE
[7] Sanjit K.Dash and G.Sasibhushana Rao, "Rob
[10] Tanmay Paul, Arnab Chakraborty, and Subhrajit Kundu, "Hybrid Shallow and Deep Learned Feature Mixture Model for Arrhythmia Classification", 978-1-5386-5135- 3/18/$31.00 2018 IEEE
[11] Sowmiya, C., and P. Sumitra. "Analytical study of heart disease diagnosis using classification techniques." In Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2017 IEEE International Conference, pp. 1-5. IEEE
[12] I Ketut Agung Enriko, Muhammad Suryanegara, Dadang Gunawan al, "Heart Disease Diagnosis System with k-Nearest Neighbors Method Using Real Clinical Medical Records", 4th International Conference, June 2018
[13] Martin Gjoreski, Anton Gradisˇek, Matjazˇ Gams, Monika Simjanoska, Ana Peterlin, Gregor Poglajen et al, "Chronic Heart Failure Detection from Heart Sounds Using a Stack of Machine-Learning Classifiers", 13th International IEEE Conference on Intelligent Environments, 2017.
[14] Sushmita Manikandan, "Heart Attack Prediction System", International Conference on Energy, Communication, Data Analytics and Soft Computing, ICCET 2017
[15] Tahira Mahboob, Rida Irfan, Bazelah Ghaffar, "Evaluating Ensemble Prediction of Coronary Heart Disease using Receiver Operating Characteristics", IEEE Internet Technologies and Application(ITA 2017)