International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-9 Issue-1, May 2020

Machine Learning for Accurate Prediction of Cardiac Arrhythmia

Rashmitha H.R, Sumana M

Abstract: ***Cardiac Arrhythmia is a state within the heart that is caused due to irregular waveforms generated from sinoatrial node. Around 17.3 million people die due o cardiac arrhythmia as indicated by World Health Organization (WHO), the kind of disruptions that is caused by sinoatrial is easily captured in Electrocardiography (ECG) readings; it records in all the disruptions and makes a record in form of images, waveforms, numerical data and categorical data. The noisy data's collected during a patient examination is recorded in form of a special character to prompt the missing data. With different set of distinct patients having different classes of arrhythmia the ECG easily records in all the arrhythmia class as Y dependent variable's that is used to pass the collected data from the ECG to the proposed system in the research study, which give's in an architectural model for detecting arrhythmia with considering a combination of Machine Learning Techniques. Random Forest is mainly used in for feature extraction for the dataset that is trained and tested followed by passing the updated dataset to a combination of different Machine Learning Techniques in order to provide accurate training and testing accuracy results from the dataset received. The use of the proposed model is in hospitals that have huge amount of dataset, with recursive training and testing of the model with the right Machine Learning Algorithm for huge amount of dataset it yields results fast in a short span of time, that can help save several life forms in a very short period of time.***

Keywords : Arrhythmia, Prediction, ECG, ML Classifiers, Accuracy.

I. INTRODUCTION

Cardiac Arrhythmia is a type of heart condition that is caused due to irregular waveforms with in the heart chambers when it receives electric signals from sinoatrial in waveform which can be too fast, slow and irregular [1]. The signals that lead to arrhythmia are caused by P100 signal which enters the sinus node in the heart to cause Cardiac arrests. The normal heart cycle involves relaxation of heart muscles with refilling it with blood, later leading to contraction in heart to pump the blood throughout the body [2].

Manuscript received on April 02, 2020.
Revised Manuscript received on April 15, 2020.
Manuscript published on May 30, 2020.

Rashmitha H.R, M.Tech, Software Engineering, M S Ramaiah Institute of Technology (MSRIT), Bangaluru.

Dr. Sumana M, Associate Professor, Department of Information Science and Engineering, M S Ramaiah Institute of Technology (MSRIT).

Arrhythmia appears to affect a human being at any time of period, while most of the sudden deaths are due to ventricular arrhythmia that leads to 80% death in the overall world. Remaining 18% of the population ends up with the common arrhythmia that is cardiovascular diseases that are treated via medications and treatments [1].

Overall according to WHO that is by World Health Organization survey around 17.3 million people end up with cardiac arrhythmia due to life style changes, due to mental health issues, excessive intake of toxins, alcohol abuse, diabetes etc. Arrhythmia is said to be affected to the people who tend to have a previous heart attack or arrest.

Arrhythmia includes three main groups they are the extra beats within the heart, supraventricular tachycardias, ventricular arrhythmias and bradyarrhythmis that in turn have many sub-groups within. Most of the arrhythmia has no prior symptoms to classify which group the arrhythmia belongs to, but when symptoms are present it is verified from the pulse caused with each heart beat and even palpitations. Most of the symptoms are not serious but is treatable, but few lead to heatstroke, heart failure or even sudden death.

ECG device is a popular way of recording in all the bioelectrical impulses from inside the heart and also a common way to detect arrhythmia patterns with the distinct distribution of heartbeats in form of PQRST components. With the use of Machine Learning techniques in the field of Medical Science it is well known to give in faster results with accuracy, with the implementation of algorithms such as Support Vector Machines, K-Nearest Neighbor, Random Forest, Naïve Bayes and many more, a model is built that is trained and tested repeatedly in order for the model to learn from its previous mistakes and give in accurate results.

II. LITERATURE SURVEY

Based on the interest of study in the field of medical science combined with machine learning algorithms the following give in brief study: Jeevan K. Pant and Sridhar Krishnan gave in a telemonitoring simulation based on the QRS complexity got from the ECG, with the feature selection implementation to the biosensors via wireless connection using the sum of absolute difference and maximum of absolute difference, gave in 7.4% reduction in computational time during the simulation process for the QRS complexity in Arrhythmia with process time of 7 millisecond [3].

Halil İbrahim BÜLBÜL, Neşe USTA and Musa YILDIZ gave in a methodology for correcting the diagnosis that is got from the ECG diagnosis, with the implementation of Back Propagation with Multi Layer classifiers and Support Vector Machine classification algorithms gave in a improvised version in time calculation and the standard classification method used during ECG diagnosis [4].

Susmita Ray gave in a research study of algorithms that can be used in the any field, with the different types of approach's that can be used to yield the best approach among all for once desired interest in fields, the review is based on many different approaches that are also the main highlights in the field of machine learning, used by superiors also stating each of the algorithms merits and demerits [5]. Prajwal Shimpi, Sanskruti Shah, Maitri Shroff and Anand Godbole gave in machine learning algorithm implementation for the detection of Arrhythmia that uses four different classifiers involving Principal Component Analysis, Random Forest, Support Vector Machine and K-Nearest Neighbor algorithm. Each of the classifiers used was cross-validated with another classifier to obtain the best trained and tested model with the end results of 91.2% for Support Vector Classifier for the classification of arrhythmia [6].

Jinkwon Kim, Hangsik Shin, Yonwook Lee, Myoungho Lee obtained dataset from MIT-BIH database with implementation of Extreme Learning Algorithm (ELM) for classification of arrhythmia that gives is in fast accuracy, they classified the heart beat got in from the ECG to 7 different classes including the normal heart beat; with the implementation of ELM they acquired 97.45% accuracy in 2.423 seconds [7]. Minh Tuan Nguyen and Kim Kiseon obtained dataset from MIT-BIH database, gave in the methodology of using convolution neural networks as a feature extraction technique for the prediction in sudden cardiac arrests that are caused due to sudden shock waveforms within the heart, with the features extracted from the convolution (CV) neural network implementation of 5-fold CV is carried out that gave in end results of 99.02% accuracy [8].

Miguel Caballero, Grace M Mirsky obtained manual collection of dataset in USA and Europe that consisted of arrhythmia alarms for reducing false accusation of arrhythmia, with the implementation of decision trees a training model of 750 multi-parameter ECG recordings were obtained in two phases that gave in 59.39 and 65.79 accuracy [9].

III. DATASET AND MACHINE LEARNING ALGORITHMS

A. ECG: Electrocardiogram

Electrocardiogram is a device that is used from 18th century by the doctors to understand the beatings that is got from the heart in form of ECG's, to also get familiar with the functionalities performed within the heart when it receives signals and to understand the physiological performance that is carried out by the heart [10]. ECG records in all the required biometrical pulses got in from the heart from its distinct pathological structure that has been evolved over years. Recordings are got in combinations of sampling frequency in Hz with duration in terms of minutes to hours. To detect arrhythmia patterns the distinct distribution of heartbeats is captured in form of PQRST components [11].

In the current research the dataset is obtained from UC Irvine Machine Learning Repository with 298 attributes and 209 consisting of non-linear dataset. The different parts in ECG dataset are as shown in Fig. 1 with, P waveform refers to the contraction of muscle in the Atrial node in the heart, QRS refers to the complexity got to refill in blood to the heart chambers that is towards the narrow end point in left and right ventricles of the heart, T refers to the relaxation of ventricles [12]. With the use of raw data procured from the ECG readings Machine Learning Algorithms is used in order to get accurate results in few seconds.

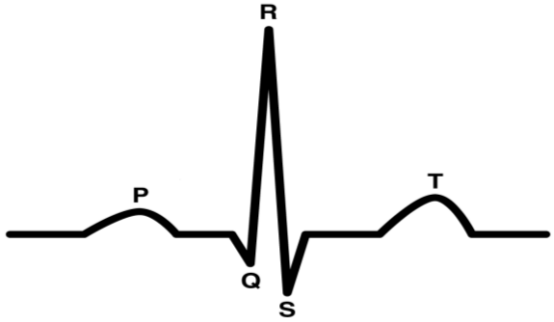

Fig. 1.ECG waveform with PQRST component

B. Machine Learning Algorithms

An overall brief review of the machine learning algorithms is stated; Machine Learning is known to be a scientific study that deals with the way a model can be trained based on past experience in terms of inferences relying on patters, following a proper set of instructions. It allows a computer the ability to build a model from the algorithm that is passed as input in order to provide accurate, sensitive, specified results and also used for forecasting purposes majorly in the field of medical science. Machine Learning's are built on the datasets that are acquired from the real world that are used to eliminate all the noisy data and get a transformed data that is used to train the model and build a mathematical model. The finally obtained trained model is tested; in order to see the best suited algorithm implementation [13].

Some of the popular classification algorithms that are used in all of fields include Random Forest, Neural Networks, Decision Trees, Support Vector Machines, Nearest Neighbor, Linear Classifiers and Principal Component Analysis. Based on the selected algorithms for implementation the feature extraction is done using a classification algorithm in order to obtain a much improvised version of dataset that yields high accuracy and prediction [1]. With the use of this sort of Machine Learning Algorithm one can easily detect the presence of arrhythmia in a patient during the early stages. The main intent of this research study is to develop a model which yields in high testing and training accuracy in results.

IV. PROPOSED MODEL

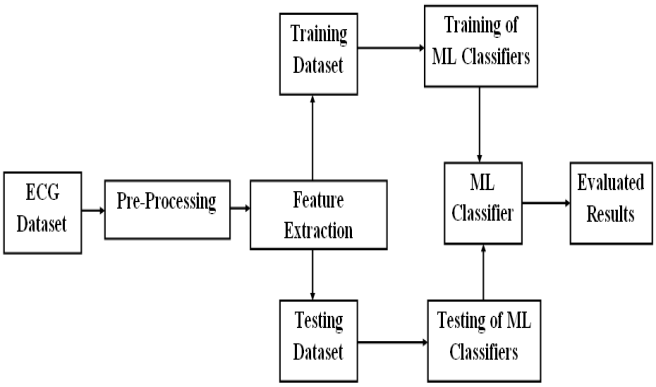

Fig. 2.Proposed System

The proposed system in this research study is as shown in Fig. 2, the raw unprocessed ECG dataset's are collected in form of images, waveforms, numerical data and categorical data and passed to the system that consist of noisy data represented by a special character and stored in columns or rows. For the removal of unwanted noisy data Pre-Processing is performed to the collected dataset. Pre-Processing is a process that is carried out in order to transform crude dataset in to readable format [14]. Some of the popularly known techniques for de-noising are the use of filters, clustering and pandas. As an end result of Pre-Processing a new transformed dataset is sent in to the feature extraction process.

Feature Extraction is carried out in order to reduce the amount of data given for analysis process without having to lose important and relevant data. Feature Extraction is carried out using a classifier algorithm that is, Random Forest that gives in reduced datasets for analysis. The new dataset got for analysis is sent in for training and testing that's taken in ratio of 80:20%. With consequent training and testing of the model each time it led to a better model that is implemented on another classification algorithm as K-nearest neighbor to see the accurate results that is obtained [15]. As a comparative study many algorithms are implemented after testing and training the model in order to get the best classifier for precise accuracy prediction.

With the use one or more Machine Learning (ML) classifiers a transformed dataset is obtained which is trained and tested based on which, the best suited algorithm for the prediction of arrhythmia is selected. As an end result it yields in training accuracy and testing accuracy. The Accuracy that is AC to the model is evaluated and obtained as in (1) where

TP is True Positive
TN is True Negative
FP is False Positive
FN is False Negative

$$\text{Accuracy (AC)} = \frac{\text{Number of Correct Prediction}}{\text{Total Number of Prediction}} \quad (1)$$

$$\text{Accuracy (AC)} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

V. RESULTS

Analysis is made for the raw dataset collected from ECG device initially included Pre-Processing for the dataset, the result is passed for feature extraction where Random Forest machine learning approach is used, it compresses the dataset information without having to lose any data and filters based on the features selected and dataset cleaned improved to 84% as shown in Table I, the dataset obtained from UC Irvine Machine Learning Repository is at a size of 796MB with 369 patient samples. The software used for this process is Spyder and language used is Pyton.

Table-I: Analysis

Approaches	Data Cleaning	Random Forest
Accuracy	79%	84%

VI. CONCLUSION

Cardiac Arrhythmia a severe type of condition in the heart, that causes many sorts of symptoms which is treatable and non-treatable. The proposed research states the use of different classification algorithms in order to identify Cardiac Arrhythmia in early stages. The model gives in high prediction accuracy with the use of ML classifiers, with consequent training and testing the model with different classifiers the best classifier for accuracy prediction is chosen. The proposed model is used with the right combination of pre-processing techniques with the ML classifier in order to give in the right technique for prediction in Cardiac Arrhythmia. With this research study many number of human species can be saved with the early detection of Cardiac Arrhythmia.

REFERENCES

1. https://en.wikipedia.org/wiki/Arrhythmia
2. https://en.wikipedia.org/wiki/Cardiac_cycle
3. Jeevan K. Pant and Sridhar Krishnan "Efficient Compressive Sensing of ECG Segments Based on MachineLearning for QRS-Based Arrhythmia Detection," *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Orlando, FL, 2016, pp. 4731-4734.
4. 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.
5. Susmita Ray "A Quick Review of Machine Learning Algorithms", 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), 978-1-7281-0211-5/19/$31.00 2019 ©IEEE.
6. Prajwal Shimpi, Sanskruti Shah, Maitri Shroff and Anand Godbole "A Machine Learning Approach for the Classification of Cardiac Arrhythmia", IEEE 2017 International Conference on Computing Methodologies and Communication (ICCMC), 978-1-5090-4890-8/17/$31.00 ©2017 IEEE.
7. Jinkwon Kim, Hangsik Shin, Yonwook Lee, Myoungho Lee "Algorithm For Classifying Arrhythmia Using Extreme Learning Machine and Principal Component Analysis", 29th Annual International Conference of the IEEE EMBS Cité Internationale, Lyon, France, 1-4244-0788-5/07/$20.00 ©2007 IEEE.
8. Minh Tuan Nguyen and Kim Kiseon "Feature Learning Using Convolution Neural Network for Cardiac Arrest Detection," *2018 International Conference on Smart Green Technology in Electrical and Information Systems (ICSGTEIS)*, Bali, Indonesia, 2018, pp. 39-42.

9. Miguel Caballero, Grace M Mirsky "Reduction of False Cardiac Arrhythmia Alarms through the use of Machine Learning Techniques", Computing in Cardiology 2015; 42:1169-1172.
10. Haresh M. Nanarkar1, Prof. Pramila M. Chawan2 " A Survey on Classification and identification of Arrhythmia using Machine Learning techniques", International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 10 | Oct 2018.
11. Ivaylo Ivanov Christov, Tatyana Dimitrova Neycheva and Rositza Todorova Raikova "ECG-noise Removal from EMG-signal by Subtraction of Hybrid Template of Averaged PQRS-T Intervals" 2019, DOI: 10.1109/ET.2019.8878620.
12. Jeevan K. Pant and Sridhar Krishnan "Efficient Compressive Sensing of ECG Segments Based on Machine Learning for QRS-Based Arrhythmia Detection" 2016 DOI: 10.1109/EMBC.2016.7591784.
13. Simon Haykin, "Neural networks and Learning Machines", 3rd Edition, PHI.
14. https://en.wikipedia.org/wiki/Data_pre-processing
15. https://en.wikipedia.org/wiki/Training,_validation,_and_test_sets

AUTHORS PROFILE

Rashmitha H.R is currently pursuing her M.Tech degree in Software Engineering fourth semester from M S Ramaiah Institute of Technology (MSRIT), Bangaluru. She has completed her B.E in Information Science and Engineering from Vivekananda Institute of Technology, Bengaluru. Currently, she is working as an intern in Philips Innovation Campus Bengaluru, India. Has been an active participant in National Level paper presentation contest on Make in India project held in November, 2016 and won Award of Excellence. Is an active participant for the conferences held for paper publishing. Her research interests include Machine Learning, Data Science, Image Processing and Cloud Computing.

Dr. Sumana M is currently working as an associate professor in the department of Information Science and Engineering, M S Ramaiah Institute of Technology (MSRIT). She has completed her PhD in Computer Science and Engineering from Manipal University in 2017. She is a researcher and involved in projects related to data science, data analytics and deep learning. Her research interests include Data Mining, Data Science, Data Analytics, Machine Learning and Deep Learning. She is an IEEE and ISTE member. She has published papers in reputed conferences and scopus indexed journals. She is an active reviewer in several conferences and journals.