**Cardio Vascular analysis using machine learning to predict early heart diseases**

**ABSTRACT:**

Myocardial infarction is a common and dangerous heart disease nowadays among all ages of people across the globe. It is also termed a heart attack. This is caused due to the physical activities, and food habits of people leading to the weakness of heart muscles. Thus weakness of heart muscles will lead to blockage in heart veins and the oxygen flow gets reduced, this might lead to death. In the existing approach, there are several techniques for identifying the heart abnormalities such as blood tetestsecho, ECG, etc. An ECG-based analysis is the most common and effective approach. But all the above-listed approaches are inspected manually, invoking manual dependency, risk of human error, time taken to process, and difficulty to inspect. In this paper, our proposal addresses the above existing approach challenges by integrating machine learning-based techniques for analyzing the user ECG inputs. In our experimental result, we use a convolutional neural network (CNN) to analyze the ECG data and identify early myocardial infarction with promising results. This project uses the Jupyter tool with python script for analysis.

***Keywords: Electrocardiography, CNN, Machine learning***

**INTRODUCTION**

A regular blood supply is required for the cardiac muscle cells to make their tasks effectively and efficiently. The coronary arteries in the heart are responsible for the supply of oxygen to the heart muscle. The weakness of the valves leads to the closure of a coronary artery, which reduces the amount of blood entering the heart. This is a portion of the myocardium that is interrupted due to a lack of blood [1]. If there is not enough blood flow and it’s not fixed in time, it will lead to the death of a person. This condition is referred to as a myocardial infarction. An arterial occlusion that can affect the normal functioning of the heart, is called a heart attack. A heart attack is an indirect sign of a heart attack, and, therefore, is called a silent heart attack. The Records show that 72% of people die as a result of a calm heart. The early detection of MI is important in today's world, ensuring proper handling and saving other people's lives. It will be located in the anterior, lateral, posterior, or lower part of the left ventricle. They can be detected by electrocardiography (ECG). This is one of the easiest and most widely used technology. An ECG records the electrical performance of the heart, in the form of signals [2].

ECG reads electrical signals in the heart that trigger the heart muscle and the blood filling the action. Twelve electrodes are attached to the skin in the chest area, arms, and legs, and for the detection of these pulses from a variety of perspectives on the INTERFACE of the waves. The ECG segments of the heart-wave and P-wave, QRS wave, and T-wave [3]. If there is a violation of the activity of the heart, this is reflected in the ECG signal. The ECG provides information on the occurrence of myocardial infarction (MI) in the event of a change in the height of the ST segment, and an incorrect response of the Q Wave and inverted T-wave. 12 different types of ECG signals are used by the CNN models to assess whether an employee is exposed or not.

The differences in the ECG signal can be detected by a cardiologist. However, there are a variety of issues in the visual interpretation of the signals. ECG signals are low in amplitude and duration so the analysis can take a long time, and inter-personal variability occurs. These drawbacks can be overcome by the use of a computer-diagnostic system for automated detection of myocardial infarction on an EKG. In deep learning CNN-based approach to working with the pass-through of architecture will be used for the detection of myocardial infarction with the use of a standard data analysis tool to data. It provides a quick, efficient, and promising analysis [4-5].

**RELATED WORKS:**

[6] Arrhythmia discovery approaches can be roughly classified based on time and frequency domain. The time-based features are assorted as those grounded-on time arena or commonness arena features. The time arena features are normally concluded out of intervals between the PQRST waves in the ECG signal, but this kind of segmentation process doesn’t provide reliable results. And also, the detection of waves except the R-peaks are very complex to extract and analyze. Thus, analyzing the RR intervals based on knowledge would help identify arrhythmia presence among the patients.

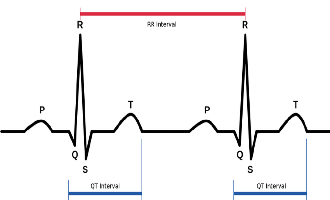
[7] The frequency domain-based features are extracted after identifying the Speaks. The wavelet-based algorithm can be proposed to extract meaningful features from the inputs. In this paper they have compared various feature extraction and classification algorithms, in this support vector machine (SVM) is considered to be an efficient classification algorithm in analyzing and identification of arrhythmia disease. SVM is efficient in detecting abnormalities from beat and rhythm level-based features.

[8] In this paper, the researchers introduced a new feature termed amplitude in diagnosing arrhythmia disease using two approaches. First, the amplitude is used in identifying the difference between heartbeat and feature extraction. Second, integration of a classifier named random forest to classify the normal and abnormality based on the features extracted. Also, many papers suggest the addition of new features in classifying heartbeat normal and abnormality in reducing false-positive results using the MIT-BIH arrhythmia database.

[9] This paper is concerned with the rank of electrocardiogram arrhythmia using a type of late­ order neural unit (HONU HONU) i.e., a QNU, with error Backpropagation in Batch optimization by Levenberg Marquardt style. The objective of the paper is to present a system that uses the classifier to help the croakers in the recognition of ECG arrhythmias and to set the performance of the QNU for ECG arrhythmia rank.

[10] The paper has encouraged us to do probation that consists of distinguishing between several arrhythmias by using deep neural network algorithms parallel as multilayer perceptron (MLP MLP) and difficulty neural network (CNN CNN). The ECG databases accessible atPhysioBank.com andkaggle.com were used for training, testing, and voucher of the MLP and CNN algorithms. The proposed algorithm in this paper consists of four secret layers with weights, and propensities in MLP. It also includes a four­ layer intricacy neural network which is used to machinate ECG samples for the different classes of arrhythmia

**PROPOSED ARCHITECTURE:**



De-Sampling

PUBLIC DATASET

Pre-Processing

UNHEALTHY

HEALTHY

**Categorize the Data**

**TRAINING DATASET**

**TRAINED DATASET**

CNN Model

**Fig 1. Proposed Architecture**

**PROPOSED METHODOLOGY:**

**Data Collection:**

We have used the open (ECG) source of data from Kaggle. This dataset is pre-processed, and de-sampling techniques can be used to make it to the classifier. The data was collected through a careful look at the examination of the databases for various research groups. The data set consists of more than 100+ records obtained from more than 100+ patients, and it is a mixture of a variety of circumstances, whether it is a valve-related heart disease or coronary artery disease.

**ECG arrhythmia classifier**

In this paper, CNN is adopted as a classifier to predict heart diseases at the early stage. CNN was first introduced by LeCun. With the introduction of the CNN model, the correlation of the spatially adjacent pixels and be obtained by the use of a non-linear filter and apply multiple filters to apply, and canract local features of the image.

The 2D convolutional neural network is more suitable for extracting the spatial localization of the ECG recordings. Therefore, for the convenience of the 2D-CNN for the classification of the ECG signals, the ECG signals in the time domain, a 2D spectrograms in the time-frequency representations [11-12].

It is also responsible for the capture of both high-level and low-level meaningful features, such as edges, color, gradient, orientation, and so on. Here are the features examined in this example, by using the input filters. The Kernels can also be known as the weight of the filter. It's the number of CPU cores used in the convolution layer, which is equal to the number of the currently selected object. In the convolution layer, the connections between the pixels, are also saved with the help of mathematical operations between the image and the kernel. It can be convolution layers. The output of the convolution layer minimizes the objects.

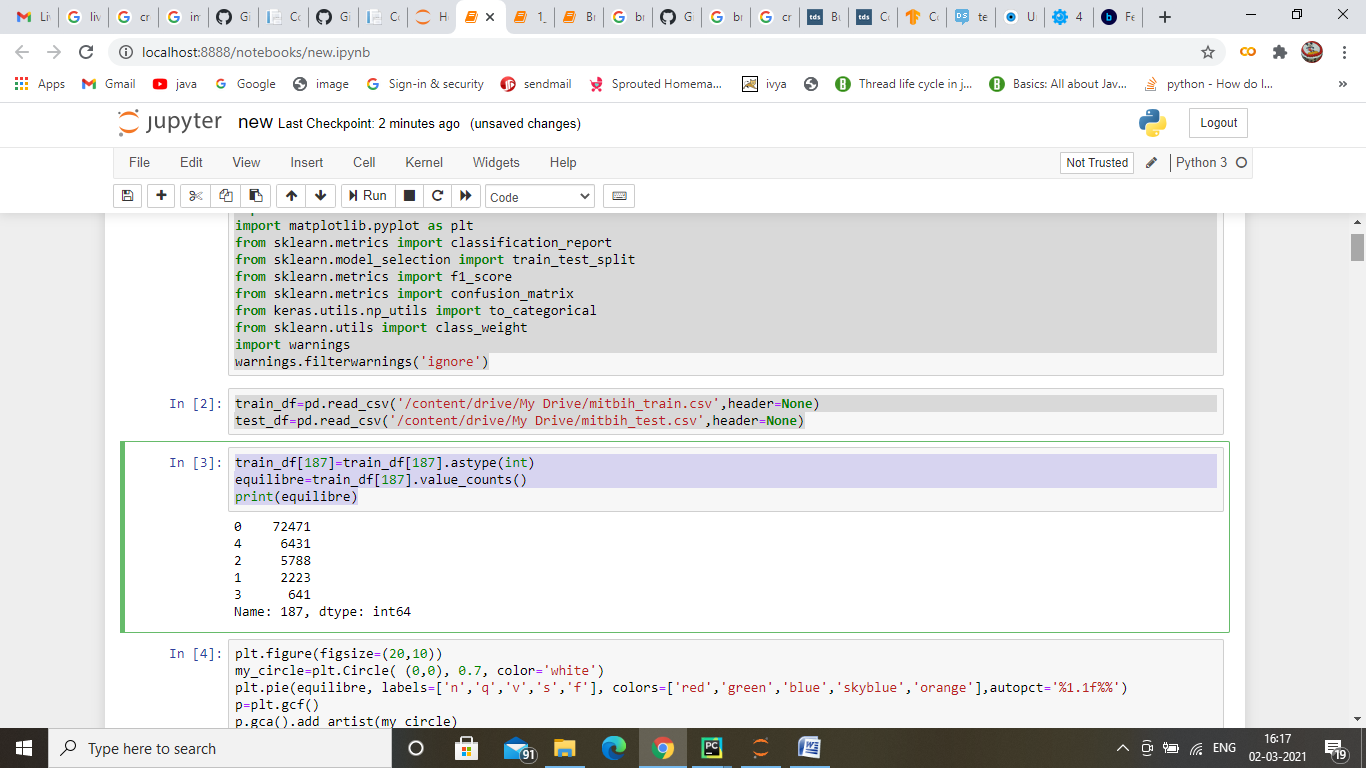
**Training the model:**

Training the model invokes two passes. Forwards pass and Backward pass. During the 1st pass, the image will go through all of the layers of the CNN model, and the final output will be calculated. In the second pass, it is determined by the value of the cost function. This process continuously takes the next input image from the training set and processes it. Based on the features extracted from the training dataset the trained model is generated.

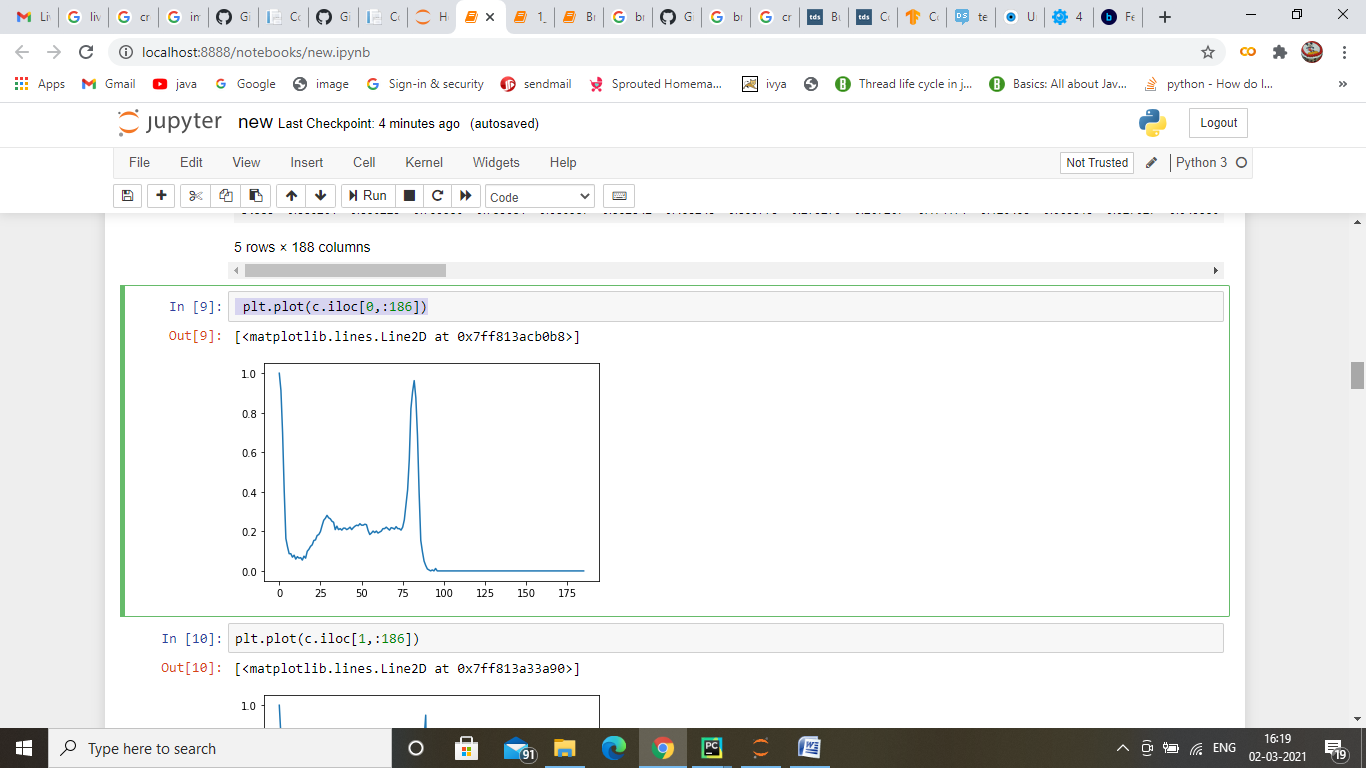
**Model**

The proposed model is based on seven numbers of convolution layers, together invoking the pooling layer as well. The image is obtained with a resolution of 128 X 128 pixels in size. For the filter, the size of each layer can vary in order, a growing population, and a sample loading of the items. Since the number of output classes in the two binary cross-entropy is used as a loss function to calculate the variation between the actual and the predicted output.

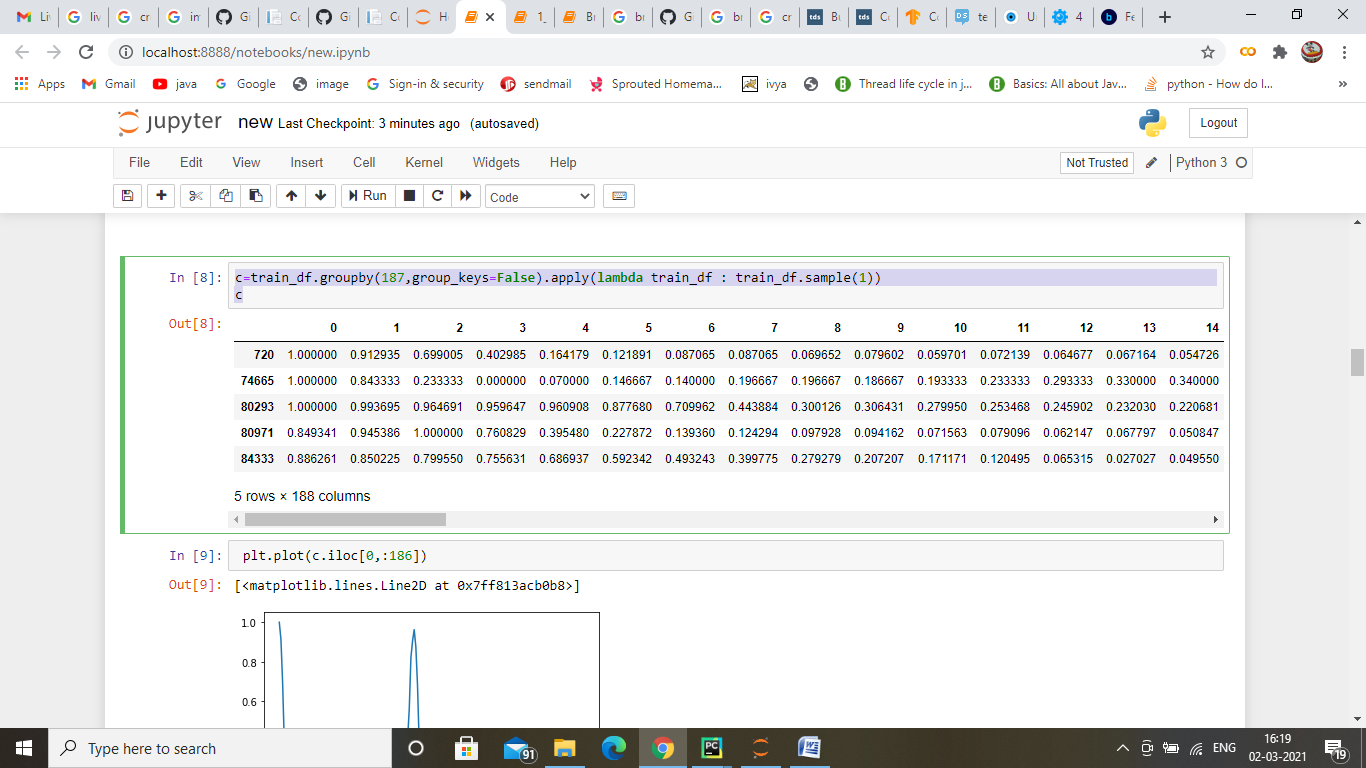
**EXPERIMENTAL RESULTS:**



**Fig 2. Jupyter Environment**



**Fig 3. Plotting based on the ECG input values**



**Fig 4. Feature extraction and analysis using CNN**

|  |  |  |
| --- | --- | --- |
| **CNN Algorithm (Recordings)** | **False Positive** | **False Negative** |
| **30** | **11** | **14** |
| **100** | **15** | **19** |

**Table 1. Performance Evaluation**

**Fig 5. Proposed System Performance Evaluation**

In the results and discussion, Fig 2 explains the Jupiter environment which can be executed online itself. All packages can be installed online itself. The programming is developed using a python script. Fig 3 explains from the input data values, that the plotting is performed to obtain the peak values. Fig 4 explains the final feature extraction of peak values and classification with the CNN model to predict the normal and abnormalities. Table 1 explains the performance of our proposed system and the output efficiency of our machine learning model. Fig 5 explains the graphical representation of the performance of the proposed classifier model.

**CONCLUSION:**

In this paper, we propose a machine learning-based technique to predict heart diseases at the early stage. In this paper, we have used a public repository dataset of 100+ recordings from 100+ different patients for data analysis. We have used CNN for classification. Based on the features and conditions, the trained model is been generated. This proposed methodology renders promising results. For experimental results, we have used python and Jupyter tools for data analysis and prediction.

**REFERENCE:**

1. World Health Organization. Cardiovascular Diseases (CVDs). Available online: http://www.who.int/ media centre/factsheets/fs317/en/ (accessed on 10 September 2017).

2. Shi, H.; Wang, H.; Zhang, F.; Huang, Y.; Zhao, L.; Liu, C. Inter-patient heartbeat classification based on region feature extraction and ensemble classifier. Biomed. Signal Process. Control 2019, 51, 97–105.

3. Kiranyaz, S.; Ince, T.; Gabbouj, M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. IEEE Trans. Biomed. Eng. 2016, 63, 664–675.

4. Minami, K.; Nakajima, H.; Toyoshima, T. Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network. IEEE Trans. Biomed. Eng. 1999, 46, 179–185.

5. Inan, O.T.; Giovangrandi, L.; Kovacs, G.T. Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features. IEEE Trans. Biomed. Eng. 2006, 53, 2507–2515.

6. Acharya, U.R.; Sudarshan, V.K.; Koh, J.E.; Martis, R.J.; Tan, J.H.; Oh, S.L.; Muhammad, A.; Hagiwara, Y.; Mookiah, M.R.K.; Chua, K.P.; et al. Application of higher-order spectra for the characterization of coronary artery disease using electrocardiogram signals. Biomed. Signal Process. Control 2017, 31, 31–43.

7. Melgani, F.; Bazi, Y. Classification of electrocardiogram signals with support vector machines and particle swarm optimization. IEEE Trans. Inf. Technol. Biomed. 2008, 12, 667–677.

8. Martis, R.J.; Acharya, U.R.; Prasad, H.; Chua, C.K.; Lim, C.M.; Suri, J.S. Application of higher-order statistics for atrial arrhythmia classification. Biomed. Signal Process. Control 2013, 8, 888–900.

9. Acharya, U.R.; Fujita, H.; Adam, M.; Lih, O.S.; Hong, T.J.; Sudarshan, V.K.; Koh, J.E. Automated Characterization of Arrhythmias Using Nonlinear Features from Tachycardia ECG Beats. In Proceedings of the IEEE International Conference on Systems Man, and Cybernetics, Budapest, Hungary, 9–12 October 2016; IEEE: New York, NY, USA, 2016; pp. 533–538.

10. Ye, C.; Kumar, B.V.K.V.; Coimbra, M.T. Heartbeat classification using morphological and dynamic features of ECG signals. IEEE Trans. Biomed. Eng. 2012, 59, 2930–2941.

11. Shi, H.; Wang, H.; Huang, Y.; Zhao, L.; Qin, C.; Liu, C. A hierarchical method based on weighted extreme gradient boosting in ECG heartbeat classification. Comput. Methods Programs Biomed. 2019, 171, 1–10.

12. Acharya, U.R.; Oh, S.L.; Hagiwara, Y.; Tan, J.H.; Adam, M.; Gertsch, A.; Tan, R.S. A deep convolutional neural network model to classify heartbeats. Comput. Biol. Med. 2017, 89, 389–396.