ISSN XXXX XXXX © 2019 IJESC

Research Article Volume 9 Issue No.3

Early Stage Detection of Cardiac Arrhythmia by Image Recognition Algorithms on Tensorflow

Monisha. N[1], Kavya. B. R[2], Hajeera Khatun[3], Sheba Jebakani[4]
BE Student[1, 2, 3], Assistant Professor[4]
Department of ISE
Atria Institute of Technology, Karnataka, India

Abstract:
Cardiac Arrhythmia is a condition where a person suffers from an irregular or abnormal heart rhythm. It is due to the malfunction in the electrical impulses within the heart that coordinate how it beats. As a result, the heart beats too fast, too slowly, or irregularly. The rhythm of the heart is controlled by a node at the top of the heart, called the sinus node, which triggers an electrical signal that travels through the heart causing the heart to beat, pumping blood around the body. Excess electrical activity in the top or bottom of the heart means that the heart doesn't pump efficiently. The most common symptoms of Arrhythmia include shortness of breath, fainting, an unexpected loss of heart function and unconsciousness that leads to death within minutes unless the person receives emergency medical treatment to restart the heart. So, it's vital to know about and understand the condition, what danger signs to look out for and how to diagnose it early.

1. INTRODUCTION

Automatic heart sound classification has promising potential to accurately detect heart diseases such as Cardiac. It could be used in non-clinical environments such as patient's residence by medical personnel as a quick heart pathology screening technique or at places with poor financial conditions to support healthcare. The focus of this project is to classify whether the patient has "normal" or "abnormal" heart sound from the Phonocardiogram (PCG) or heartbeat recordings to quickly identify patients who would require further diagnosis. This is a supervised learning problem since we already know if the heart sound in training dataset is normal or abnormal. The basic idea is to convert each heart sound recording(wav file) to a spectrogram image and train a Convolutional Neural Network over those images. Then given a new PCG recording, we will be able to classify it as normal or abnormal.

2. LITERATURE SURVEY

The research paper as in [1] showed that using multi scale infor-mation in wavelets it is very easy for the characterization of the ECG signal and the QRS complex. > Senhadi(1995) made a com-parison of tha wavelet transforms for the recognization of cardiac patterns. > Spline wavelet (6 levels) and complex wavelet (10 levels) provide Daubechies decompositions where choosing of the wavelet family as well as the selection of the analyzing function into these families. > Even though D6 algorithm is more complex and has a slightly higher computation but it picks detail that is missed by the wavelet algorithm, which is simpler than the D6 algorithm (proved by Amara Graps 1995). > Zong and Jiang (1998) introduced a fuzzy reasoning approach for the ECG signal and rhythm identification and classification. > Based on the fuzzy logic Sugiura (1998) found a new technique for cardiac arrhythmia detection that can seper-ate NSR from VF and a method for discrimination of ventricular arrhythmia. > Acharya (2003) by using the heart rate variability as the initial signal extracted a certain parameters which were used for the fuzzy equivalance for the classificatiuon of four cardiac arrhythmias. > This was 90% accurate over all the cases > To the proposed ANF Kannathal (2005) used three non-linear parameters as inputs for determination of heart abnormalities. The research paper as in [2] chazal used wavelet coefficient for classification ECG. > The performance of the wavelet was accurate comparing to other methods but it suffers from many problems such as consuming time, static window size and it depends on nominal features that make the analysis and classification very difficult. >Tsipoura (2004) proposed method for classifying arrhythmias based on R-intervals used by the rules that were determined by the medical doctors. > Even though this method has high performance and easy to understand but has many problems such as time consuming, doesn't classify all cardiac arrhythmias. > Iske (2007) classified the ECG signal by extracting the features in time domain and frequency domain and then he used support vector machine (SVM). > The problem in this method, SVM maps input vectors to the high dimensional space and makes a problem in ECG analysis because if the dimensional space is high the error rate will be high. The research paper as in [3] analyzed in the time domain thus corresponding arrhythmias are determined by using ANN, around 95%result is achieved for identification of arrhythmia. > The classification of arrhythmias and compared it with standard data storage and diagnosis different disease. >The SVM, Random-Forest and KNN are used for the classification of cardiac arrhythmia. >The results were compared and the accuracy of each of the algorithm is calculated. > The filter part deals with feature selection from the cardiac arrhythmia dataset of the UCI machine learning repository. The research paper as in [4] is an approach that is commonly deployed for effective ECG analysis. > The researches aimed at increasing the accuracy rate for ECG beats,for which MIT-BIH databases of 92 patients records are extracted. > The proposed sys-tem resulted an accuracy of 99% with acquisition of high clustering performance. In The research paper as in [5] The ECG classification is done using two classification techniques 1.Supervised learning. 2. Unsupervised learning. > It has the feature selection methods for choosing a number of features among original features. > The advantage of feature selection is reduction in time and cost of feature acquisition. > Feature selection is also helpful in improving classifier accuracy, provided that noisy, irrelevant features are eliminated. The research paper as in [6] they take the ECG signals and converts the analogue signal to a digital waveform. > Then the system will extracted 8 beats from each ECG signal samples of 2223 samples per second and classifies the patients beats. > The next step is that the signal wills pre-processing the data which was denoising of loaded raw ECG signal. > Later the system will extract only three features from the signal; QRS complex duration, and also the RR interval both normal and the one averaged over 8 beats.> And also these features were later used by ANN classifiers such as Naive Bayes and Multi-class SVM to predict the class of the arrhythmia. > Finally the results were compared with the datasets and the accuracy of each of the algorithm was calculated. The research paper as in [7] describes feature of extracting the methods using higher order statistics (HOS) of wavelet packet decomposition (WPD). > The coefficients used for making automatic heartbeat recognition. > The wavelet package coefficients (WPC) is been calculated for the different type of ECG beat. Later the higher order of WPC is been derived. > Finally, the calculated feature data set is used for input to a classifier, and it is based on k-NN algorithm. >They have used the MIT-BIH arrhythmia database to obtain the ECG records. > The patients heartbeats in the arrhythmia is taken as the data set which are grouped into five main heartbeat classes. > The accuracy of this paper study is measured by average sensitivity of 90%, average selectivity of 92% and average specificity of 98%. The research paper as in [8] presents a method for analyzing electrocardiogram (ECG) signal. > Firstly, they extract the features to classify the heart beats according to different arrhythmias. > The Data was taken from 40 records of the MIT-BIH arrhythmia database. > They choose learning dataset which was obtained from a twenty records of data set that was manually classified. > Fast Fourier transforms were used to identify the peak levels in the ECG signal and later Neural Networks are applied to identify the diseases. >They used Levenberg Marquardt Back-Propagation algorithm to train the network.

The research paper as in [9] used Fuzzy logic and Artificial Neural Network (FANN), Precise Electrocardiogram (ECG) classification to analyse patient's conditions. > Mainly ECG signal are classification of dataset is performed using various pulses for example v1, v2, v3, v4, v5, v6 etc. > P-Wave, PR-Interval, QRS-Interval, ST-Interval, T-Wave for analysis of each Input pulse used to train the network. >Finally output of the network gives weight factors of each signals to create a data set. > Data sets were organized by clusters. Later these cluster data sets are analysed by Adaptive Resonance theory. The research paper as in [10] uses discrete wavelet transform (DWT) for extraction of ECG signals obtained from MIT-BIH Arrhythmia Database. > The Machine Learning Technique, Probabilis-tic Neural Network (PNN) was used to classify four types of heart beats that consist of PVC, LBBB, RBBB and Normal.

3. TENSORFLOW

Tensorflow is a computational framework for building machine learning models. TensorFlow provides a variety of different toolk-it's that allow you to construct models at your preferred level of abstraction. You can use lower-level APIs to build models by defining a series of mathematical operations. Alternatively, you can use higher-level APIs (like tf.estimator) to specify predefined architectures, such as linear regressors or neural networks. The following figure shows the current hierarchy of TensorFlow toolkits:

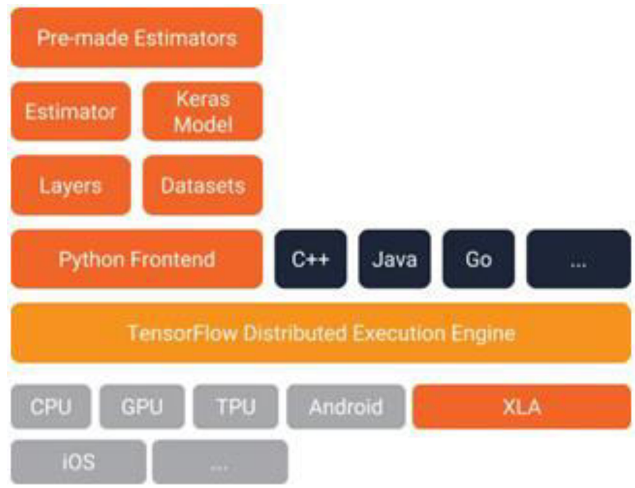

Figure.1. Tensorflow Architecture

For this particular project the architecture that is used around Tensorflow is as in Figure 2

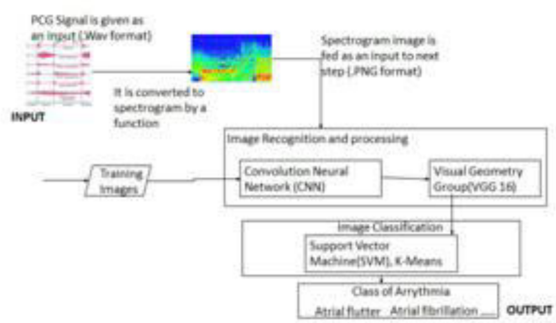

Figure.2. Neural Network Tensorflow as per our project Detection of Cardiac Arrhythmia

4. DATASET

Tensorflow as per the above diagram helps us to classify the images, but we need a large dataset of images to train the neural network for it to be able to classify. The dataset used for this training is available freely as part of the PhysioNet / Computing in Cardiology Challenge 2016 which focuses on automatic classification of normal /abnormal phonocardiogram (PCG) recording. The dataset has 4,430 recordings taken from 1,072 subjects, collected from both healthy subjects and patients with a variety of conditions such as heart valve disease. Along with clean heart sounds, the dataset also contains some noisy recordings. The samples have been obtained from both normal subjects and pathological patients, providing a variety of signal sources. The training data consists of PCG signals of varying length, any-where between 5s to just over 120s all sampled at 2000 Hz and are provided in .wav format. Each recording contains only one PCG lead. Each recording has been labeled as normal or abnormal in a separate file(.hea format). As expected, the distribution of data is highly imbalanced and therefore, various data augmentation techniques have been implemented. Since the dataset is very large in size, it was decided to work with a smaller subset. There-fore, the training set is chosen to be about 1000 labeled recordings and the validation set to be 200 recordings.

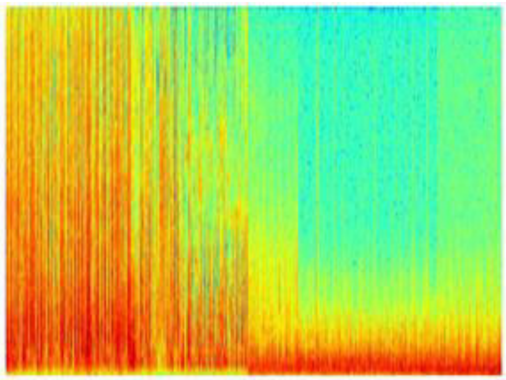

Figure .3. Example of abnormal PCG

5. ALGORITHMS USED FOR IMPLEMENTATION

1. Decision Tree Algorithm

Decision tree is one of the most popular machine learning algorithms used all along. Decision trees are used for both classification and regression problems, this story we talk about classification. Decision tress often mimic the human level thinking so its so simple to understand the data and make some good interpretations. A decision tree is a tree where each node represents a feature(attribute), each link(branch) represents a decision(rule) and each leaf represents an outcome(categorical or continues value). The whole idea

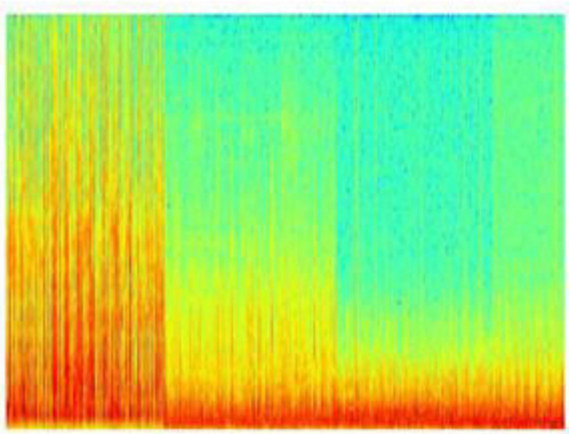

Figure .4. Example of normal PCG

Is to create a tree like this for the entire data and process a single outcome at every leaf(or minimize the error in every leaf).

Decision Tree Algorithm Pseudocode

1. Place the best attribute of the dataset at the root of the tree.
2. Split the training set into subsets. Subsets should be made in such a way that each subset contains data with the same value for an attribute.
3. Repeat step 1 and step 2 on each subset until you find leaf nodes in all the branches of the tree.

2. KNN Algorithm

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure. We can implement a KNN model by following the below steps:

1. Load the datase
2. Initialize the value of k
3. For getting the predicted class, iterate from 1 to total number of training data points, calculate the distance between test data and each row of training data. Here we will use Euclidean distance as our distance metric since it's the most popular method. The other metrics that can be used are Chebyshev, cosine, etc.
4. Sort the calculated distances in ascending order based on distance values
5. Get top k rows from the sorted array
6. Get the most frequent class of these rows
7. Return the predicted class

6. INTERFACE DESCRIPTION AND DJANGO ARCHITECTURE

Django follows a Model-View-Controller (MVC) architecture, which is split up into three different parts:
The Model is the logical data structure behind the entire applica-tion and is represented by a database (generally relational databases such as MySql, Postgres). The View is the user interface — what you see in your browser when you visit a website. These are represented by HTML/CSS/Javascript files.The Controller is the middleman

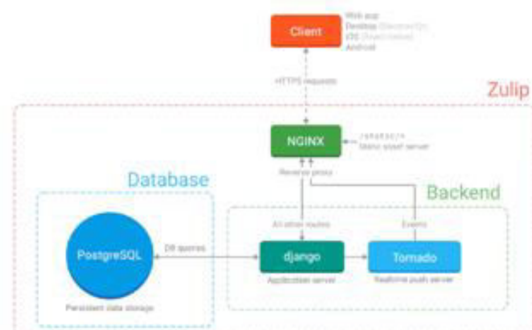

Figure .5. Django Architecture

That connects the view and model together, meaning that it is the one passing data from the model to the view. With MVC, your application will revolve around the model—either displaying it or manipulating it. So say a user will enter a URL in their browser, that request will go through the internet protocols, to your server, which will call Django. Django will then process the given URL path, and if it matches an URL path you have explicitly stated, it will call the Controller, which will then perform a certain action, such as get an entry from your Model(database) and then render a View(ie: JSON text, HTML/CSS/JavaScript Web page).

7. EVALUATION METRIC

The metrics to determine how well the model performs on the entire dataset is logarithmic loss which is also named as ‘categorial cross entropy’. This can be described as negative the log likelihood of the model given each observation is chosen independently from a distribution that places the predicted probability mass on the corresponding class, for each observation. In our case, each recording is already labeled with one true class and for each one, a set of predicted probabilities is considered an outcome. Here N is the number of images in the test set, M is the number of image class labels i.e 2, log is the natural logarithm, Yi,j is 1 if observation belongs to class and 0 otherwise, and Pi,j is the predicted probability that observation belongs to given class. Since the dataset is imbalanced i.e number of normal samples is greater than number of abnormal samples in the training dataset, accuracy is not only the metric we will consider.

8. CONCLUSION

Since the dataset is imbalanced i.e number of normal samples is greater than number of abnormal samples in the training dataset, accuracy is not only the metric we will consider. So the model will be evaluated on ‘precision’ which can be described as the ratio of correct positive predictions made out of the total positive predictions made We also calculate another metric known as ‘fbeta score’ which is given by the weighted average of precision and recall. Thus, it is possible to obtain almost near precision in this classification.

9. REFERENCES

[1]. Cardiac Arrhythmia Classification Using Fuzzy Classifiers - Mrs. B.Anuradha, V.C.Veera Reddy [2008]

[2]. Classification of Cardiac Arrhythmias using Machine learning techniques based on ECG Signal Matching - Makki Akasha Babikier, Mohammed Izzeldin, Ibrahim Musa Ishag, Dong Gyu Lee,Keun Ho Ryu

[3]. A Literature Review on Arrhythmia Analysis of ECG Signal - S.T.Sanamdikar , Dr.S.T.Hamde , Dr.V.G.Asutkar

[4]. A Survey on various Machine Learning Approaches for ECG Analysis - C.K Roopa, B.S Harish

[5]. Arrhythmia classification using SVM with selected features-N Kohli, NK Verma

[6]. Study and Literature Survey for Classification of Cardiac Arrhythmia with respect to ECG and HRV Signals - Mr. Vyankatesh S. Thorat1, Dr. Suresh D. Shirbahadurkar2, Mrs. Vaishali V. Thorat3

[7]. Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients. - Kutlu Y, and Kuntalp D, 2012

[8]. Cardiac arrhythmias detection in an ECG beat signal using fast fourier transform and artificial neural network - Gothwal H, Kedawat S, and Kumar R, 2011

[9]. Survey on Cardiac Arrhythmia Techniques in ECG Signal - Srinivas N, VinayBabu A, and Rajak M.D, 2013

[10]. Electrocardiogram Beat Classification using Probabilistic Neural Network - Banupriya C.V, and Karpagavalli.S, 2014