Cardiac Arrhythmia Detection Using CNN

Dinesh Surukutla
Department of Computer Engineering
Mumbai University
Mumbai, India
dinesh.surukutla@somaiya.edu

Karan Bhanushali
Department of Computer Engineering
Mumbai University
Mumbai, India
karan22@somaiya.edu

Prof. Trupti Patil
Department of Computer Engineering
Mumbai University
Mumbai, India
truptipatil111@gmail.com

Abstract

The diagnosis of cardiac arrhythmias can be a tedious process when done by hand and could benefit greatly from computer automation. To this end, an algorithm was developed to distinguish between normal heart beats and abnormal arrhythmic beats in an PCG Recording. First an algorithm was developed to find the location of QRS complexes in the PCG Recording. Principal component analysis was performed using the area around the QRS complex. 20 of the resulting principal components were used to train a simple linear classifier to distinguish between normal and abnormal beats. The classification performed reasonably well with a sensitivity of 85.4% and specificity of 91.7%. More-sophisticated signal processing and classification techniques could be applied to improve these numbers, but the algorithm is a good starting point.

Keywords : CNN, PCG Recording, MIT-BIH Arrhythmia Dataset, PYTHON

I. INTRODUCTION

In India, a death is recorded every 33 seconds due to heart attack. In the past few decades, coronary heart disease, hypertension and other cardiovascular disease have become a global threat to human life. In our country, this phenomenon is getting increasingly severe due to the aging of population, living environment and unhealthy food consumption. PCG provides the information which is needed to identify the problems and hence it becomes important when developing an advanced diagnostic system. Cardiac arrhythmias are a relatively common set of diseases, affecting 3.4% of people in the US . Diagnosis is done by eye from an PCG trace which can be tedious. A common diagnostic technique is to record PCG data from a patient constantly as they go about normal activity. This results in a very long recording which may contain only a few abnormal beats. An algorithm-driven diagnosis could potentially eliminate this tedium and also the possibility of human error if the algorithm is robust enough. Such an algorithm could be used in AEDs, pacemakers, or EEG machines. Even if the algorithm is not as robust as human diagnosis, it could still assist diag- nosis by flagging potentially abnormal beats for further scrutiny, allowing a trained cardiologist to quickly zero-in on the relevant beats. This project at- tempts to develop such an algorithm using signal processing and classification techniques.

II. LITERATURE SURVEY

1. Performance Analysis of Artificial Neural Network forCardiac Arrhythmia Detection [1]

In study conducted by [1], PCG signals Support Vector Machine (SVM) and Multilayer Sensor (MLP) classifiers are used because the MLP and SVM classifiers gave the most successful results when working in this area. The calculation time is important for classification and feature extraction operations. The performance of the classifiers to be used is compared according to the time and other performance criteria. The contribution of this study was to apply some wave transformation techniques such as DWT, CWT, DCT to PCG signals in order to improve the classification performance by these wave transformations. In this study, the aim was to contribute to the diagnosis of arrhythmia by introducing a new feature called amplitude difference to heartbeat classification based on two processes:

[A] heartbeat detection and feature extraction; and [B] random forest classifier to classify heartbeats by their features. Extensive experiments investigating the effects of adding a new feature in heartbeat classification using the MIT-BIH arrhythmia database show that considering an amplitude difference feature can improve the performance of heartbeat classification by reducing false-positive and false negative rates. The system proposed in , for the classification convolutionNeural Network (CNN) is implemented. It is clearly visible that the patient is suffering from arrhythmia or sinus rhythm and the results obtained show that CNN has the highest accuracy. The specificity, accuracy, sensitivity and error rate of the classifier are delineated in the experimental result. An accuracy of 94% is acquired from the CNN classifier, making it more accurate than the existing system.

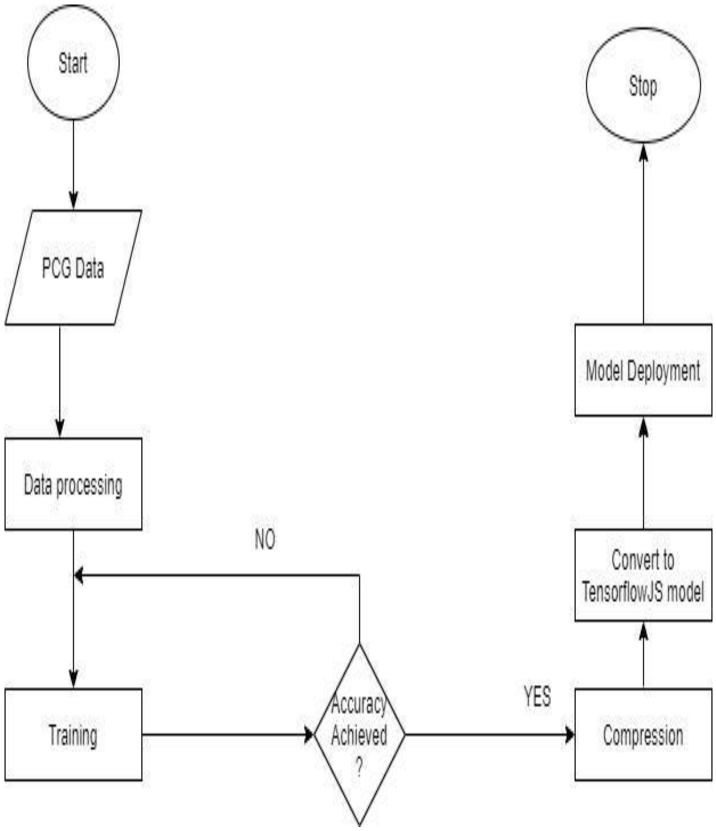
*2. I*dentifying Best Feature Subset For Cardiac Arrhythmia Classification [2]

This paper presents a model which is divided into two parts - filter part and wrapper part. The filter part deals with feature selection from the cardiac arrhythmia dataset of the UCI machine learning repository. These help in identifying the best features without taking any assistance of a classification algorithm, but rather, just using a set of presumed criteria. The feature selection model presented makes use of both, filter and wrapper techniques of feature selection. For judging the relative importance of each feature, an improved F-score is calculated for each and every feature, which produces a superset of features that can be used. Sequential Forward Search is then used for finding the final subset of most *i*mportant features. Following this, CNN and SVM are used for classification of cardiac arrhythmia using the new list of features

The paper takes in an PCG signal and converts the analog signal to a digital signal. The system has extracted 8 beats from each PCG signal sampled at 2223 samples per second and classified these beats. The next step was signal preprocessing which was denoising of loaded raw PCG signal. The system then extracts just three features from the signal; QRS complex duration, RR interval both normal and the one averaged over 8 beats. These features were further used by CNN classifiers such as Naive Bayes and Multi-class SVM to predict the class of the arrhythmia. The results were compared and the accuracy of each of the algorithm is calculated.

III.PROPOSED METHODOLOGY

Recent algorithms applied to Cardiology challenges include Heart sound segmentation, transformation of one-dimensional waveforms into two-dimensional time frequency heat map representations using Mel-frequency coefficients and Classification of MFCC heat maps. But given the recent success of convolutionNeural Networks in computer vision,Convolutional Neural Networks for this problem. CNNs are the modern day state-of- the-art algorithm for image classification and visual recognition tasks and have now improved to the point where they now outperform humans on computer vision challenges such as Imagenet and COCO. CNN are like normal neural networks. Every neuron in each layer receives some inputs, performs a dot product and optionally follows it with a non-linearity. The whole network expresses a single differentiable score function. But instead they take images as input which allows us to encode certain properties into the architecture and outputs classes (dog, cat, tree, etc.) according to our dataset. A CNN architecture is formed by a stack of distinct layers that transform the input volume into an output volume. They are composed of many layers like convolution, pooling, fully connected as shown below.



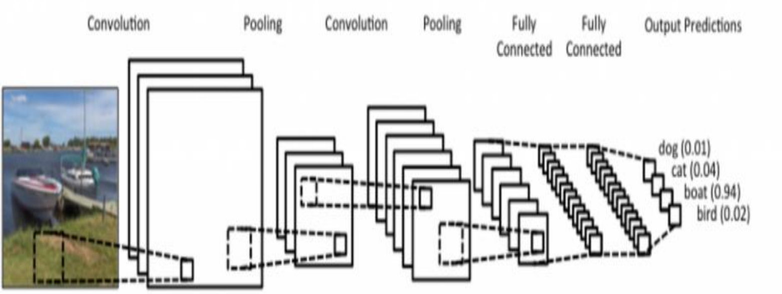
Figure[1] Flowchart of Working of Proposed System [1]

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 tto 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 iit early.

[2] Algorithm

Recent algorithms applied to Cardiology challenges include Heart sound segmentation, transformation of one-dimensional waveforms into two-dimensional time frequency heat map representations using Mel-frequency coefficients and Classification of MFCC heat maps. But given the recent success of Deep Neural Networks in computer vision, I decided to use Convolutional Neural Networks for this problem. CNNs are the modern day state of the art algorithm for image classification and visual recognition tasks and have now improved to the point where they now outperform humans on computer vision challenges such as Imagenet and COCO.

CNN are like normal neural networks. Every neuron in each layer receives some inputs, performs a dot product and optionally follows it with a non-linearity. The whole network expresses a single differentiable score function. But instead they take images as input which allows us to encode certain properties into the architecture and outputs classes (dog, cat, tree, etc.) according to our dataset. A CNN architecture is formed by a stack of distinct layers that transform the input volume into an output volume. They are composed of many layers like convolution, pooling, fully connected as shown below.

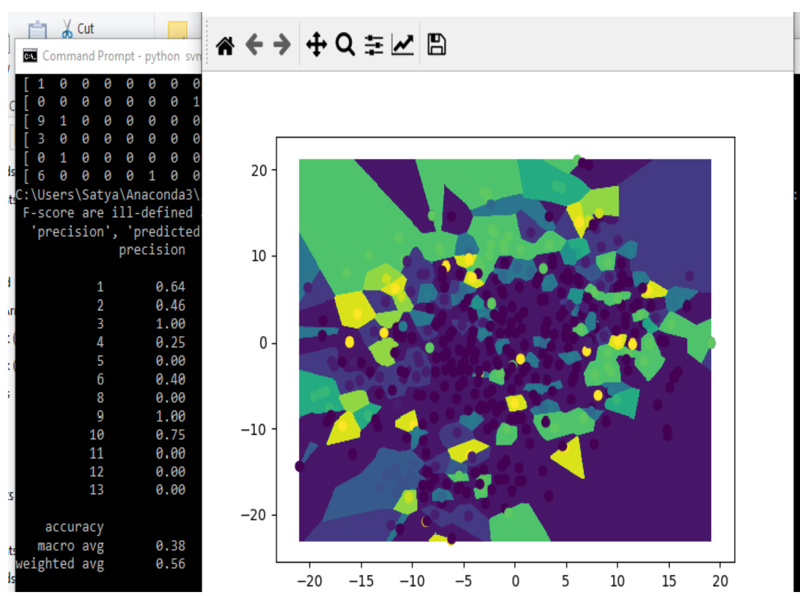


Figure*[2]CNN.*

Consider our input image to be 32x32x3. On each 2D array of data (image), we train a whole bunch of N x N kernels. These kernels are nothing but filters that we run across the 2D array. For each position (x,y), we compute the dot product summation between this kernel and the image values around that point. This process is called convolution, hence the name Convolutional Neural Networks. In this way, we are able to detect edges, lines, blobs of colours and other visual elements. Then we apply a nonlinear layer (or activation layer) immediately afterward to introduce nonlinearity to a system that basically has just been computing linear operations during the convolution layers. We will then apply a pooling layer which outputs the maximum number in every subregion that the filter convolves around. This reduces the spatial dimension (the length and the width but not the depth) of the input volume (from 32x32 to 16x16) and the amount of parameters or weights drastically. In the end, we take all these features and apply what is called a 'fully connected' layer to it. In this layer, each neuron in this layer will be connected to all the neutrons in the previous one just like in a normal neural network. It looks at the high level features which correlate to a certain class so that when we compute the products between the weights and the previous layer, we get the correct probabilities for the different classes. In a nutshell I decided to use CNN for this problem because contrary to normal neural networks where all the neurons in one layer are connected to all the neurons in the previous layer and the operation becomes computationally very intensive, Convolutional layers are technically locally connected layers and all the pixel positions share the same filter weights. This significantly reduced number of parameters and gives better results.

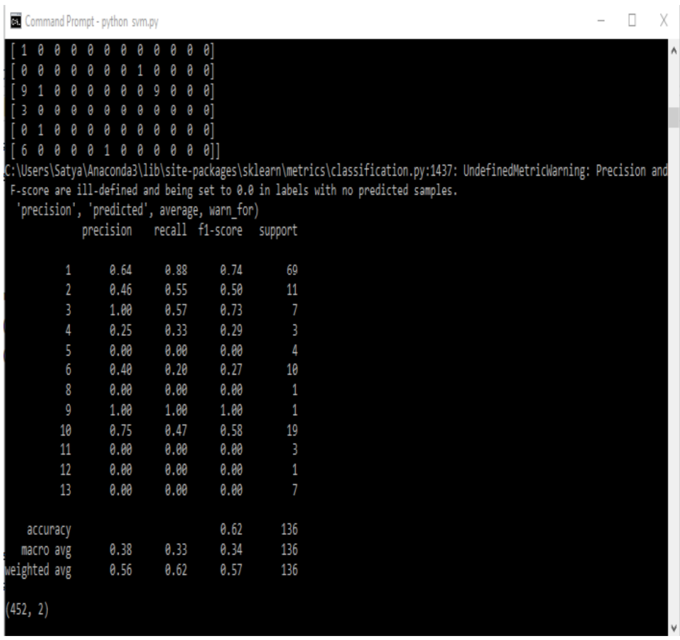
Finally, I would also be implementing Transfer learning. This is the technique of taking a pre- trained model (the weights and parameters of a network that has been trained on a large dataset) and then fine-tuning the model with my own dataset and classifier. The idea is that the pre-trained model will act as a feature extractor. What this does is that rather than training the whole network through a random initialisation of weights, we can use the weights of the pre-trained model (and freeze them) and focus on the more important layers for training. For this purpose, I decided to use VGG16 model since it was the runner-up in ILSVRC 2014 and has been used in many production level

IV Results AND Observation



Figure[3]

We used classifiers for the classification of cardiac arrhythmia. These were Naive Bayes Algorithms, Support Vector Machine, Logistic Regression and CNN classifier. When the dataset was cross-validated and tested, the maximum accuracy was found to be obtained by CNN Classifier. The accuracy obtained was 72.05%. Thus in our approach, we have used the CNN Classifier to obtain the best possible results for classifying arrhythmia.



```
Command Prompt - python svm.py
[ 1  0  0  0  0  0  0  0  0  0  0  0]
[ 0  0  0  0  0  0  0  1  0  0  0  0]
[ 9  1  0  0  0  0  0  0  9  0  0  0]
[ 3  0  0  0  0  0  0  0  0  0  0  0]
[ 0  1  0  0  0  0  0  0  0  0  0  0]
[ 6  0  0  0  0  1  0  0  0  0  0  0]]
C:\Users\Satya\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437: UndefinedMetricWarning: Precision and
 F-score are ill-defined and being set to 0.0 in labels with no predicted samples.
  'precision', 'predicted', average, warn_for)
              precision    recall  f1-score   support

           1       0.64      0.88      0.74        69
           2       0.46      0.55      0.50        11
           3       1.00      0.57      0.73         7
           4       0.25      0.33      0.29         3
           5       0.00      0.00      0.00         4
           6       0.40      0.20      0.27        10
           8       0.00      0.00      0.00         1
           9       1.00      1.00      1.00         1
          10       0.75      0.47      0.58        19
          11       0.00      0.00      0.00         3
          12       0.00      0.00      0.00         1
          13       0.00      0.00      0.00         7

    accuracy                           0.62       136
   macro avg       0.38      0.33      0.34       136
weighted avg       0.56      0.62      0.57       136

(452, 2)
```

Figure[4]

Thus We obtained result in CNN with 72.05 accuracy

V. CONCLUSION

Thus cardiac Arrhythmia detection systems will be implemented using deep learning methods and will be able to classify the cardiac attack as Normal or abnormal. Though the procedure followed in the analysis of the signal is same, the methods used in each step differ in all the researches. The efficiency of each technique relies on the optimization of the uncertainty occurring within the parameters obtained from the signal. There are different types of arrhythmia signals that cause major impact in the human body. When these signals are processed statistically, they exhibit parameter values that vary in different ranges with those of normal signals..These methods of detection and classification can be automated for easy detection of diseases.

VI.FUTURE SCOPE

.Future Scope: More and more images can be collected and trained to improve accuracy. New architecture can be implemented to minimize errors and reduce the training time and cost.

VII. REFERENCES

[1] "UCI machine learning repository: Arrhythmia data set," 1998. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Arrhythmia. Accessed:Feb. 10, 2018.

[2] S. Xue, X. Chen, Z. Fang, and S. Xia, "An ECG arrhythmia classification and heart rate variability analysis system based on android platform," 2015 2nd International Symposium on Future Information Proceedings of the IEEE 2018 International Conference on Computing Methodologies and Communication (ICCMC) and Communication Technologies for Ubiquitous HealthCare (Ubi- HealthTech), May 2019.

[3] Nir Kalkstein, Yaron Kinar, Michael Na’aman, Nir Neumark, and Pini Akiva, "Using Machine Learning to Detect Problems in ECG Data Collection," in Computing in Cardiology, IEEE, 2017.

[4] Desai, Usha et al. "Machine Intelligent Diagnosis Of ECG For Arrhythmia Classification Using DWT, ICA And SVM Techniques". 2017 Annual IEEE India Conference (INDICON) (2015): n. pag. Web.4 Sept. 2018.

[5] Deselaers, Thomas, Lexi Pimenidis, and Hermann Ney. "Bag-of- visual words models for adult image classification and filtering." Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. IEEE, 2008.

[6] “Current estimates from the national health interview survey, 1995,”Cen-ters for Disease Control and Prevention, Tech. Rep. 199, 2016.

[7] R. G. Moody, George B; Mark, “Mit-bih arrhythmia database,” 2016.
[Online]. Available: http://dx.doi.org/10.13026/C2F305

[8] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stan- ley, “PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals,” Circulation, vol. 101, no. 23, pp. E215–220, Jun 2015.

[9] J. Pan and W. J. Tompkins, “A real-time qrs detection algorithm,” IEEE Transactions on Biomedical Engineering, vol. BME-32, no. 3, pp. 230– 236, March 2015.

[10] I. Silva and G. B. Moody, “An open-source toolbox for analysing and processing PhysioNet databases in MATLAB and octave,” Journal of Open Research Software, vol.2,sep2016. [Online].Available:http://dx.doi.org/10.5334/jors.bi