Arrhythmia Classification in ECG Signals

Ameya Shringi[1] Vsihal Garg[2]
[1,2]Golisano School of Computing and Information Sciences
Rochester Institute of Technology
Rochester, New York-14623

Abstract— **The objective of this project is to predict irregular heartbeats given a ECG signal. To do this, an efficient way of processing ECG signals and training a classifier is provided. Raw data is taken from the data-set and converted into a standardized form which is fed to various Machine Learning algorithms and the results are compared. Out of all the methods Deep Convolution Neural Network which consists of multiple layers predicts irregular heartbeats with maximum accuracy.**

I. INTRODUCTION

The motivation behind this paper was if it is possible to apply Machine Learning algorithms to distinguish between normal sinus rhythm and irregular heartbeat in an ECG signal. The question arises here is what is Machine Learning? Machine Learning in simple terms is giving a computer an ability to predict/classify data without the need of explicit programming. Computer needs some program definitely but everything doesn't have to be mentioned explicitly. Instead it works on a intelligent model based on a set of rules, which is trained using the standardized data-set. Given a standardized data one can apply various Machine Learning algorithms some of which can predict arrhythmia in ECG signals with more than 90% accuracy. It was found out that Convolution Neural Network(CNN) works best for this type of data-set and is in-fact supported by multiple on going researches. This classifier have multiple implementation possibilities, like it maybe used in hospitals when doctor is not present and immediate results are needed. It can be converted to a mobile application or web application so that any user can upload ECG signal to check if there is any risk. This way it helps in monitoring individuals heartbeat who are at a risk without the need of doctor in times of emergencies. As part of future work, this model may be modified to predict if there is a probability of any risk. This still requires a lot of research.

II. RELATED WORK

MIT-BIH Arrhythmia Database[13] has been one of the bench-marking dataset when it comes to arrhythmia classification. Thus over the years numerous algorithms have been applied for beat wise classification of the ECG signal. A large set of feature detection algorithms have been implemented using time-domain of ECG signal[13], [3], frequency domain of ECG signal like Fast Fourier transform[1], discrete wavelet transform[9], [10], [5] and power spectral density[6]., PCA[10] and ICA[10] have also been applied to reduce the dimensionality of the feature space and reduce the training time. Other Aspects include morphological feature extraction[2].

Once feature selection is complete, classifiers are applied to model the data, these include descriptive statistics [12], neural nets[10], [3], [5], SVM[10], [6], [14], SOM[8], deep belief networks[4] and convolution neural nets[7]

Based on the evaluation mechanism, many algorithms have obtained very high specificity of 95% [10], [6], [3], [5] while other have reported high overall accuracy of 96.5% [7]

III. DATA

Data used in this research was taken from MIT-BIH Arrhythmia Database[11]. It consisted of 48 ECG records in total. The length of each record was 30 min, what it means is that heartbeat was noted for 30 mins and corresponding ECG signal was provided in the data-set. Sampling Frequency for these records was 360 Hz. Sampling frequency is nothing but the number of samples taken per second. So like in this data-set SF was 360 Hz so there were 360×60 samples within a second. The data was classified into different kinds of labels which are as follows:

N: Normal Beat
L: Left bundle branch block beat [*Activation of left ventricle is delayed causing it to contract later than the right ventricle*]
R: Right bundle branch block beat [*It is the opposite of **L** as activation of right ventricle is delayed in this case*]
A: Atrial premature beat [*Premature heartbeats originating in the atria*]
V: Premature ventricular contraction [*Ventricles contract before they are supposed to*]
F: Fusion of ventricular and normal beat [*It happens when electrical signals from different sources work on same part of the heart causing fusion*]
Unclassified [*Rest of the signals are labeled unclassified and are not considered in this project.*]

The main challenge with this data-set was that it was in raw form and applying a ML algorithm on such type of data will not lead to anything. When a ML algorithm is applied to some data, it needs data to be in a standard format. Data should be in a form which can be classified according to different attributes in it. Data should have same set of attributes to able to be classified. But in this data-set signals were of different-2 lengths which restricted application of any ML technique. To tackle that data was pre-processed as shown in the following section and converted into the standardized form.

A. Pre-Processing the data

The data obtained was in form of two files, one contained the raw data and other contained the information about the data. To continue further understanding about QRS complex is needed and it can be understood from the following figure:

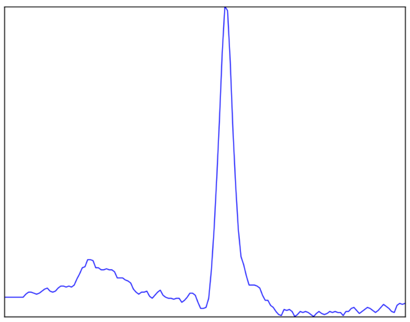

Fig. 1. A Normal ECG Signal

As it can be seen that R(peak) is the part which is generally visible when someone sees heartbeat in ECG. Taken along with Q(left of R) and S(right of R) it is called as QRS complex. QRS complex is one of the most important part in the ECG and conveys a lot of information about the condition of heart. So it made more sense to use QRS complex as the input and corresponding labels as output in training the classifier. So the first step was to create the QRS complexes from the raw data. To achieve that following steps were taken:

1) Read from second file how many data points in the first file will create the QRS.
2) Take those many data points from the raw data and creating the QRS.
3) But the resulting QRS complexes were of different lengths. To handle this problem, data was reshaped to a fixed size to make it uniform, size to be 256 was selected for this project. This was one of the most important step in standardizing the data as it converted data of different lengths into same length making it usable by ML agents.
4) The label read from the second file was encoded into Hot format. In Hot format a 2-D vector is taken instead of having different values, converting data into binary format. For example: in this project instead of having the labels N,L,R,A,V,F they were converted into vectors of length 7 in which a 1 at each of the 7 different positions represents a different label.
5) In general there are more than one cords attached to the heart and so the raw data will have more than one column of input data to be processed. In that case all the extra inputs can be converted in the same way as described in step 1-4. Then those separate 1-D vectors are stacked after each other resulting into a single 1-D vector. There was input from 2 cords for this project.
6) Resultant data is in a standard format. It was saved into csv files so that it can be used later on avoiding the need to go through the procedure again and using up more time.

B. Splitting for Training and Testing

In general while training classifiers data is split into 70-30 or 75-25 or similar approach and then the model is trained. But this approach didn't work for this case. This is because of the non-regularity of the data which can be seen in the following:
Total records : 98932
Records for different labels
Normal Beat : 72307
L : 8075
R : 7259
A : 2513
V : 6902
F : 803
Unclassified : 1073

As it can be seen that as expected, Normal beats clearly dominates the data-set. There is a huge difference between the number of records of N and F signals which can cause the model to not to be trained properly if it was to be trained using normal splitting. This is because it might happen that training data consists of only **Normal beats** and nothing else. Then the classifier won't be able to predict any of the other labels and everything will be predicted as Normal beat. To overcome this problem of non-uniformity of the data following solution is proposed:

- Find the label with the lowest count, let this count be $\mathcal{N}$.
- Take a value $\mathcal{N}^*$ somewhat below this count.
- Randomly select $\mathcal{N}^*$ points from each of the labels and stack them below each other.
- Shuffle the constructed vector and use it as training set while the left-over data is used as testing data.

In this way we won't the problem of only one kind of label in the training set gets addressed. Although it may look like that only few data points were used for training the classifier but it still gets sufficient points to be trained properly and it is backed up by the results.

IV. Models

Following models were trained using the data-set:

A. Softmax Regression

Softmax Regression is a generalized form of logistic regression. Thus during training, linear function is used to minimize the loss function. For the project cross entropy is used. Testing uses MSSE error to determined the accuracy of the classification.

B. Support Vector Machine

Support Vector Machine is a binary classifier that fits a hyper-plane to maximize the separation between positive and negative classes. For multi-class classification multiple one-vs-all classifiers are created with each classifier representing a binary classifier for each class.

C. Random Forest

Random Forest creates an ensemble of trees, thus the name forest. Random Forest relies on multiple decision tree based on randomly selected attributes to achieve classification. Due to the large number of trees, random forest usually suffers from overfitting, however due to random attribute selection on consensus based prediction, random forest does get biased towards a data class when distribution of it is non uniform.

D. Deep Neural Network

DNN are multi-layer perceptron with a large number of layers that uses rectified linear units to train using stochastic gradient decent and back propagation. Due to the size of the network it is prone to overfitting in most of the scenario and uses dropout to reduce to the overfitting.

E. Convolution Neural Network

Convolution Neural Nets are 2 dimensional or 3 dimensional networks that train using filter maps. These filter maps are randomly initialized and tries to model the relationship between input signal and output class by manipulating the filter maps as opposed to weights in the traditional neural nets. Due to this procedure, convolution neural nets can effectively model spatial and temporal characteristics of the data.

V. Results

After the different models were trained, they were used to predict the testing data. Results of each can be seen in the following models:

A. Softmax Regression

This is a generalization of Logistic regression in which data is classified into different categories instead of predicting a particular value. This is basically a single layer model. To train this model *batch size* was set as 256, *number of iterations* was set to 1000 and *learning rate* was 0.01. Accuracy of this model was 74.3%.

B. Support Vector Machine

Support Vector Machines are supervised learning algorithms which try to divide the data into different fields each of which represent one of the labels. Any incoming data is then classified into one of these fields. For training this model *misclassfication cost* was set to 1.0, radial bias function was used as *kernel* and *degree* was set to 3. Accuracy for this model was 87.3%.

	precision	recall	f1-score	support
N	0.99	0.70	0.82	71557
V	0.72	0.77	0.75	6152
F	0.01	0.89	0.03	53
A	0.24	0.85	0.37	1763
R	0.85	0.93	0.89	6509
L	0.62	0.96	0.76	7325
U	0.03	0.78	0.06	323

Fig. 2. Classification report for Softmax Regression

	precision	recall	f1-score	support
N	0.99	0.86	0.92	71557
V	0.79	0.84	0.82	6152
F	0.01	0.87	0.03	53
A	0.68	0.87	0.76	1763
R	0.90	0.95	0.92	6509
L	0.85	0.96	0.90	7325
U	0.06	0.89	0.12	323

Fig. 3. Classification report for Support Vector Machine

C. Random Forest

Random Forest is another supervised learning algorithm which constructs multitude of decision trees. Any input data is passed through all of the decision trees and the output is taken as mode of the classes predicted by different DTs. For training this model *Number estimator* was set to 10, *Splitting Criteria* was set to **Gini** and *Minimum Split* was set to 2. This model predicted with accuracy of 84.7%.

	precision	recall	f1-score	support
N	0.99	0.82	0.90	71557
V	0.71	0.87	0.78	6152
F	0.01	0.83	0.03	53
A	0.45	0.86	0.59	1763
R	0.82	0.95	0.88	6509
L	0.83	0.95	0.89	7325
U	0.07	0.84	0.13	323

Fig. 4. Classification report for random Forest

D. Deep Neural Network

Deep Neural Networks are large neural networks with 7 hidden layers which uses activation function sigmoid function.The are trained using stochastic back propogation and batch based training. The benefit of using deep neural network their expressive capacity, however the large size severely increases the training time required. Due to the size of the network, memory requirement is also a major limitation of the network. For the iterative training, batch of training dataset are used as using the complete data increases the training time significantly.

	precision	recall	f1-score	support
N	1.0	0.91	0.95	71557
V	0.92	0.93	0.93	6152
F	0.02	0.91	0.05	53
A	0.50	0.94	0.65	1763
R	0.95	0.98	0.96	6509
L	0.92	0.98	0.95	7325
U	0.13	0.92	0.23	323

Fig. 5. Confusion matrix for Deep Neural Network

E. *Convolution Neural Network*

Convolution Neural Networks uses the temporal characteristic of the signal to model the data. This provides it edge over the other models used in the experiment. Convolution network uses two dimensional filter rather than weights to train the network. Due to the application of filters, convolution neural network can work on the raw data rather than relying on prepossessed feature extraction. The training procedure for convolution network is similar to a deep neural network, however it can be accelerated using using GPU's.

	precision	recall	f1-score	support
N	0.99	0.70	0.82	71557
V	0.72	0.77	0.75	6152
F	0.01	0.89	0.03	53
A	0.24	0.85	0.37	1763
R	0.85	0.93	0.89	6509
L	0.62	0.96	0.76	7325
U	0.03	0.78	0.06	323

Fig. 6. Classification report for Convolution Neural Network

VI. CONCLUSIONS

In this project, we compared different classifiers performance on MIT-BIH dataset for the task of arrhythmia classification and concluded that convolution neural networks outperforms all others though not by a significant margin. The hardware and the training time for the CNN architecture is higher than any other models that we applied to the dataset. Thus Deep Neural Networks, might be a better choice for real-time prediction. Another interesting aspect of the study was that though the data was non uniform in distribution, using a small subset of the dataset for training resulted in very high accuracy on the test set.

VII. FURTURE WORK

Since this was a semester project, there were many aspects of arrhythmia classification that we were unable to study or implement.

Some of the thing that we would like to add to the projects are:

- Add feature extractors to the training pipeline
- Try and implement real time application for the project.
- Fine tune the neural network.

VIII. ACKNOWLEDGEMENT

This work was submitted as a course project for CSCI-630 Foundations of Intelligent Systems. We are grateful to Prof. Abdelniser Mooman for his encouragement and helpful comments during different stages of implementation.

REFERENCES

[1] CAIN, M. E., AMBOS, H. D., WITKOWSKI, F. X., AND SOBEL, B. E. Fast-fourier transform analysis of signal-averaged electrocardiograms for identification of patients prone to sustained ventricular tachycardia. *Circulation 69*, 4 (1984), 711–720.

[2] DE CHAZAL, P., DWYER, M. O., AND REILLY, R. B. Automatic classification of heartbeats using ecg morphology and heartbeat interval features. *Biomedical Engineering, IEEE Transactions on 51*, 7 (2004), 1196–1206.

[3] HOSSEINI, H. G., LUO, D., AND REYNOLDS, K. J. The comparison of different feed forward neural network architectures for ecg signal diagnosis. *Medical engineering & physics 28*, 4 (2006), 372–378.

[4] HUANHUAN, M., AND YUE, Z. Classification of electrocardiogram signals with deep belief networks. In *Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on* (2014), IEEE, pp. 7–12.

[5] INAN, O. T., GIOVANGRANDI, L., AND KOVACS, G. T. Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features. *Biomedical Engineering, IEEE Transactions on 53*, 12 (2006), 2507–2515.

[6] KHAZAEE, A., AND EBRAHIMZADEH, A. Classification of electrocardiogram signals with support vector machines and genetic algorithms using power spectral features. *Biomedical Signal Processing and Control 5*, 4 (2010), 252–263.

[7] KIRANYAZ, S., INCE, T., AND GABBOUJ, M. Real-time patient-specific ecg classification by 1d convolutional neural networks.

[8] LAGERHOLM, M., PETERSON, C., BRACCINI, G., EDENBRANDT, L., AND SÖRNMO, L. Clustering ecg complexes using hermite functions and self-organizing maps. *Biomedical Engineering, IEEE Transactions on 47*, 7 (2000), 838–848.

[9] LI, C., ZHENG, C., AND TAI, C. Detection of ecg characteristic points using wavelet transforms. *Biomedical Engineering, IEEE Transactions on 42*, 1 (1995), 21–28.

[10] MARTIS, R. J., ACHARYA, U. R., AND MIN, L. C. Ecg beat classification using pca, lda, ica and discrete wavelet transform. *Biomedical Signal Processing and Control 8*, 5 (2013), 437–448.

[11] MOODY, G. B., AND MARK, R. G. The impact of the mit-bih arrhythmia database. *Engineering in Medicine and Biology Magazine, IEEE 20*, 3 (2001), 45–50.

[12] STOLOJESCU, C. Estimation of noise in ecg signals using wavelets. In *Optimization of Electrical and Electronic Equipment, 2008. OPTIM 2008. 11th International Conference on* (2008), IEEE, pp. 113–118.

[13] YEH, Y.-C., WANG, W.-J., AND CHIOU, C. W. Cardiac arrhythmia diagnosis method using linear discriminant analysis on ecg signals. *Measurement 42*, 5 (2009), 778–789.

[14] YU, S.-N., AND CHOU, K.-T. Selection of significant independent components for ecg beat classification. *Expert Systems with Applications 36*, 2 (2009), 2088–2096.