Classification of ECG Heartbeat Arrhythmia Project

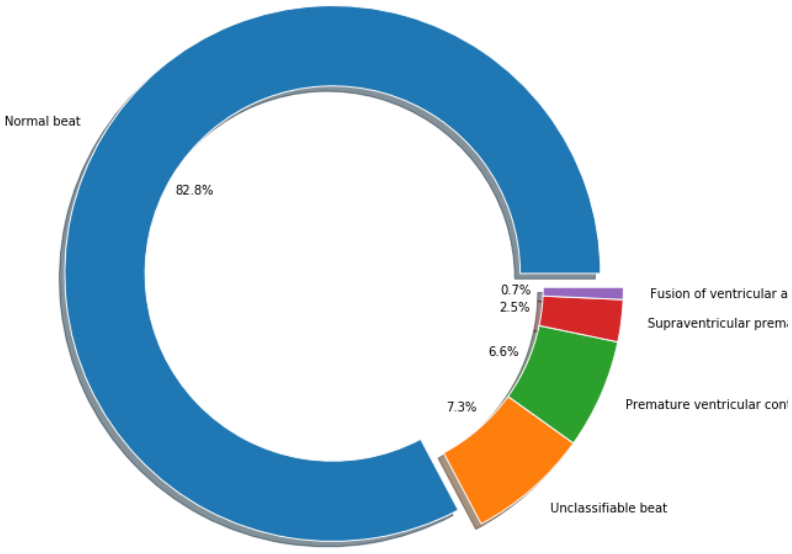
January 15, 2021

1 Introducing the Problem

We will use the the MIT-BIH Arrhythmia Dataset from Kaggle: https://www.kaggle.com/shayanfazeli/heartbeat

The dataset consists of 109446 ECG samples each one being a heartbeat which is divided in a training set consisting of 87554 samples and the rest form the test set. The problem is to build a model to classify heartbeats into different types of arrythmia. Each heartbeat is characterized as one of five, in total, different types of arrhythmia.

In this project we will deal with time series of same length, with each one representing a heartbeat. It's expected that we will deal with a classification problem of imbalanced classes as, we have the following statistical information abou the train and test set combined:



As a classification problem we will use different machine learning algorithms like SVM with and without PCA, as well as deep learning models in order to build the best model for this classification task.

2 Data Preprocessing

In order to process the data we proceeded to the following steps:

- We checked for missing values and we have found none
- We used resampling in order to overcome the imbalance of the classes

3 Feature engineering

We used neural networks and Support Vector Machines to approach the problem of classifying the differetn types of arrythmia.

- For the training and the evaluation of the neural networks we used one hot encoding to transform the labels of the hearbeats
- For the Support Vector Machines we used a grid cross validation technique for hyperparameter tuning. Also we checked the perfomance with and without PCA

4 Models

We have used three different models on this problem:

4.1 PCA - SVM :

The first model is a SVM model with an rbf kernel. Fristly we will use the model without PCA wiht the following parameters:

- C = 45
- gamma = 0.1

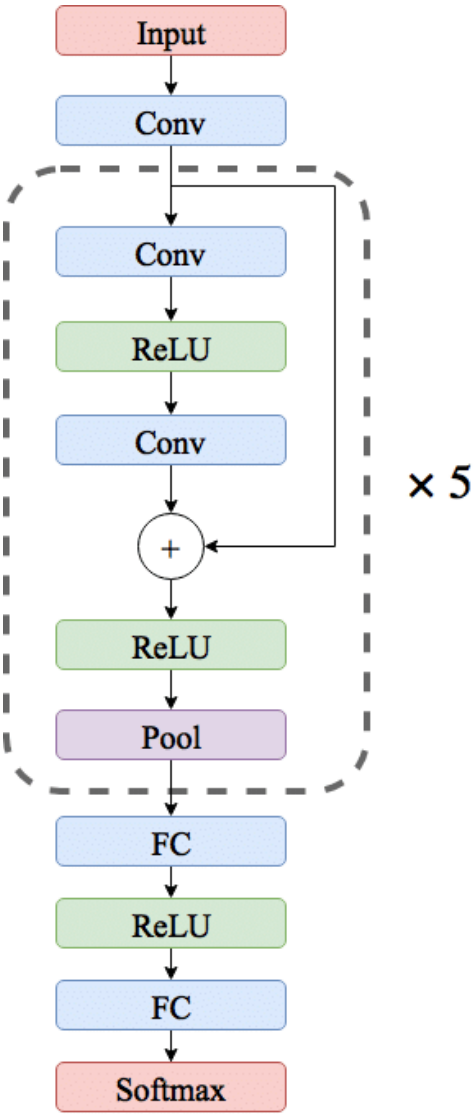
Then we will use grid cross validation for hyperparameter tuning, which are the following:

- principal components = 13
- C = 45
- gamma = 0.1

4.2 Residual Convolutional Network :

Next we used a convolutional neural network using the architecture proposed in the paper [KFS18] by Kachuee et. al.

The architecture of the net proposed in the paper can be seen below:

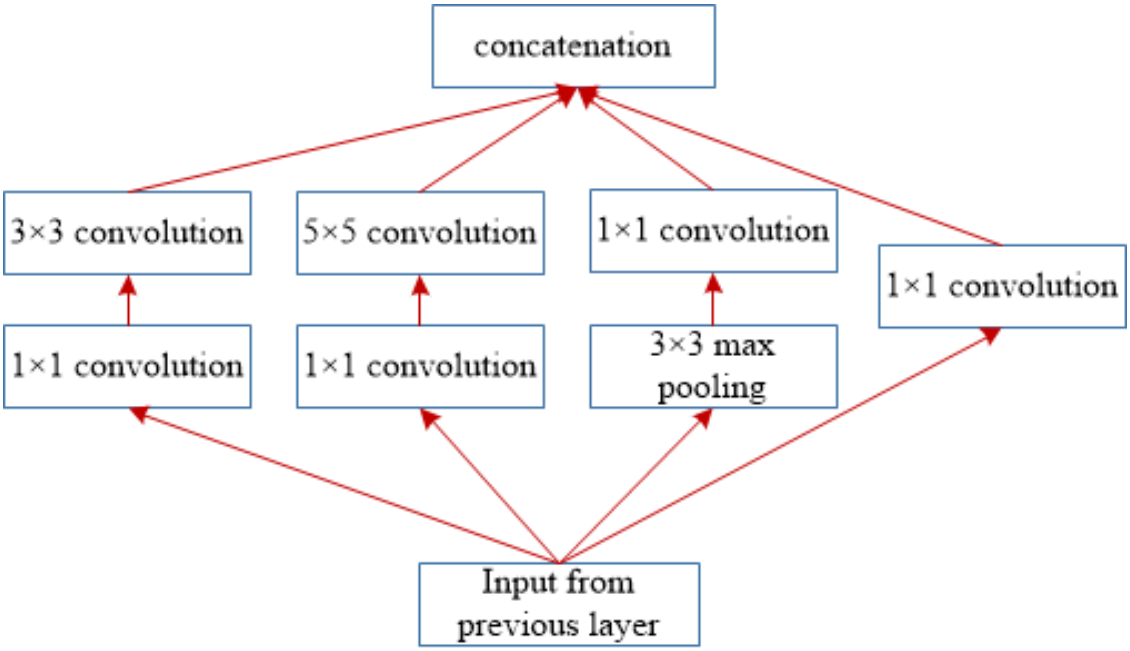


Description:
Total params: 55,013
Trainable params: 55,013
Non-trainable params: 0

We have trained the neural network for 30 epochs.

4.3 Network with inception blocks :

The last model we used is a convolutional neural network using the inception blocks architecture. The architecture of a basic inception block can be seen below:



Description:
Total params: 4,053,061
Trainable params: 4,053,061
Non-trainable params: 0

He have trained the network for 5 epochs.

5 Results

5.1 SVM :

TOTAL TIME: 8.590498952070872 minutes

BALANCED ACCURACY: 0.9120627201713548

CLASSIFICATION REPORT:

	precision	recall	f1-score	support
N	0.99	0.97	0.98	18118
S	0.58	0.82	0.68	556
V	0.91	0.94	0.93	1448
F	0.44	0.85	0.58	162
Q	0.98	0.98	0.98	1608
accuracy			0.96	21892
macro avg	0.78	0.91	0.83	21892
weighted avg	0.97	0.96	0.96	21892

5.2 PCA - SVM :

BALANCED ACCURACY: 0.9071815554960809

CLASSIFICATION REPORT:

	precision	recall	f1-score	support
N	0.99	0.89	0.94	18118
S	0.32	0.83	0.46	556
V	0.84	0.91	0.87	1448
F	0.17	0.93	0.29	162
Q	0.91	0.97	0.94	1608
accuracy			0.90	21892
macro avg	0.65	0.91	0.70	21892
weighted avg	0.95	0.90	0.92	21892

5.3 Residual Convolutional Network :

TOTAL TIME: 0.6849900325139363 minutes

BALANCED ACCURACY: 0.9228672182817697

CLASSIFICATION REPORT:

	precision	recall	f1-score	support
N	0.99	0.98	0.98	18118
S	0.61	0.84	0.70	556
V	0.94	0.93	0.94	1448
F	0.63	0.88	0.74	162
Q	0.97	0.99	0.98	1608
accuracy			0.97	21892
macro avg	0.83	0.92	0.87	21892
weighted avg	0.97	0.97	0.97	21892

5.4 Network with inception blocks :

TOTAL TIME: 12.68472870985667 minutes

BALANCED ACCURACY: 0.935212511051738

CLASSIFICATION REPORT:

	precision	recall	f1-score	support

```
              N       0.99      0.98      0.99     18118
              S       0.67      0.87      0.76       556
              V       0.93      0.96      0.94      1448
              F       0.64      0.88      0.74       162
              Q       0.98      0.99      0.98      1608

    accuracy                            0.97     21892
   macro avg       0.84      0.94      0.88     21892
weighted avg       0.98      0.97      0.98     21892
```

References

[KFS18] M. Kachuee, S. Fazeli, and M. Sarrafzadeh. "ECG Heartbeat Classification: A Deep Transferable Representation". In: *2018 IEEE International Conference on Healthcare Informatics (ICHI)*. 2018, pp. 443–444. DOI: 10.1109/ICHI.2018.00092.