Atrial Fibrillation Classification with Recurrent Neural Networks

December 8, 2019

Contents

1 Introduction

Today, with the advent of portable one-lead ECG devices, there is a significant incentive to develop algorithms that are able to detect cardiac anomalies. For example, Atrial Fibrillation (AF), defined as uncoordinated atrial activation [1], which generally results in deterioration of atrial mechanical functions, is an arrhythmia whose early detection can significantly increase the patient's chance from retaining long-term, debilitating consequences, or even death. Since the advent of Deep Learning and the development progress of GPU computing, we have found that a possible contribution to this field could be Recurrent Neural Networks. As such this paper proposes a solution based on Long-Short-Term-Memory (LSTM) Neural Networks which is able to consistently distinguish AF signals from normal sinus wave, alternate rhythms, or just noise. This is done in the context of one-lead ECG signals, with the dataset provided by Physionet's AF Classification Challenge from 2017.

Before we go directly to the results and the model proposed, we will first provide some background to the challenge and the data. Then we will provide background to the deep learning model proposed. Following the model theory, we will provide the results, along with notes regarding the experiments. Finally, we will conclude the report with final notes regarding the experiment in general, and what we can expect moving forward.

2 Data and Models

In this section, we will provide more information about the data and the models developed in the solution.

2.1 Data and Challenge

The data used for this experiment comes from the Physionet challenge, proposed in 2017. This challenge aimed at providing an incentive to the community to apply modern techniques to help the problem of AF detection and classification. The training data is composed of 8528 ECG recording that range from 9-60 seconds. Moreover, a validation set of 300 recordings is also provided, so we can evaluate the performance of the proposed model in a completely different sample set.

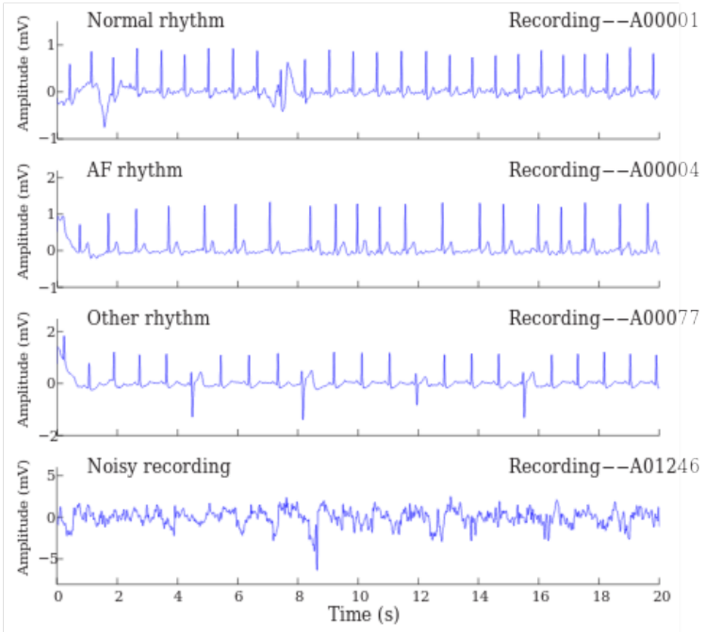

Figure 1: The different classes in the dataset and their respective waveforms.

In Figure 1, it is possible to see the differences between the different type of waveforms. By inspection, it is possible to note the general differences in frequencies of the signals, as the noise waveform presents consistently higher frequency cycles that difficult to distinguish, whereas a normal rhythm presents a consistent clear periodic signal. The classes and their respective mapping are presented in Table 1. The transformation of labels proposed was made so to facilitate the classification exercise.

Classification	Original Dataset Label	Transformed Label
Normal	N	0
AF	A	1
Other	O	2
Noisy	~	3

Table 1: Description of labels contained in the challenge dataset.

Now, it was previously mentioned that the data included ECG strips that ranged from 9 to 60 seconds in length. In addition, the each recording was provided in 300 Hz. This means that each input sample has a variable number of dimensions. In order to mitigate this issue, we padded the ending of each strip, to produce a uniform dataset of length 18256

2.2 Models

Over the course of this project, two different models were developed. In the first model, we included information about the frequency domain of the signal to the model. In the second model, we only fed the signal straight to the model.

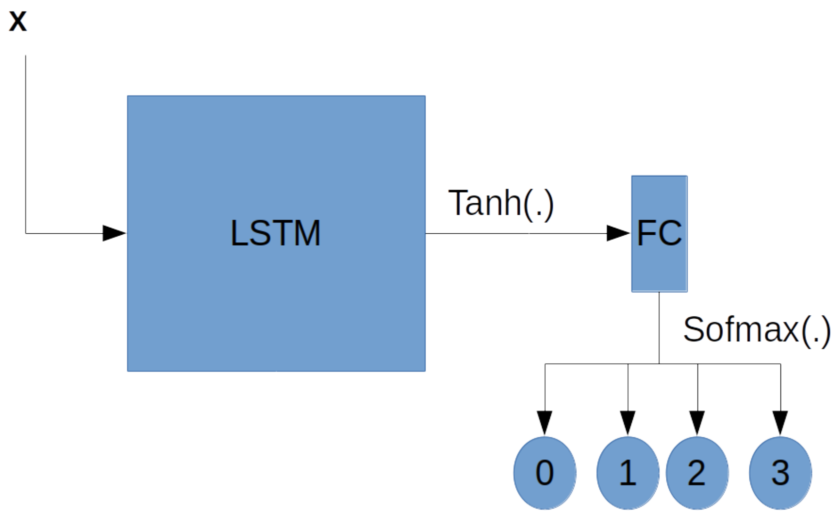

Figure 2: Basic LSTM model used

As Figure 2 shows, the basic model is an LSTM layer, followed by a fully connected layer. The LSTM reduced the dimension of the data to $\frac{1}{5}$ of the dimension of the of input. The dimension of the input varied, however, the results reported in section 3 refers to an input dimension of 9000. This means that the we considered up to 30 seconds worth of information for each ECG strip. The output of the LSTM layer was put through a hyperbolic tangent function, and later a fully connected layer. The fully connected layer reduced the dimension of the data to the required class dimension of 4.

This result was then placed through a softmax function, which then provided the probabilities that the input belonged to a specific class.

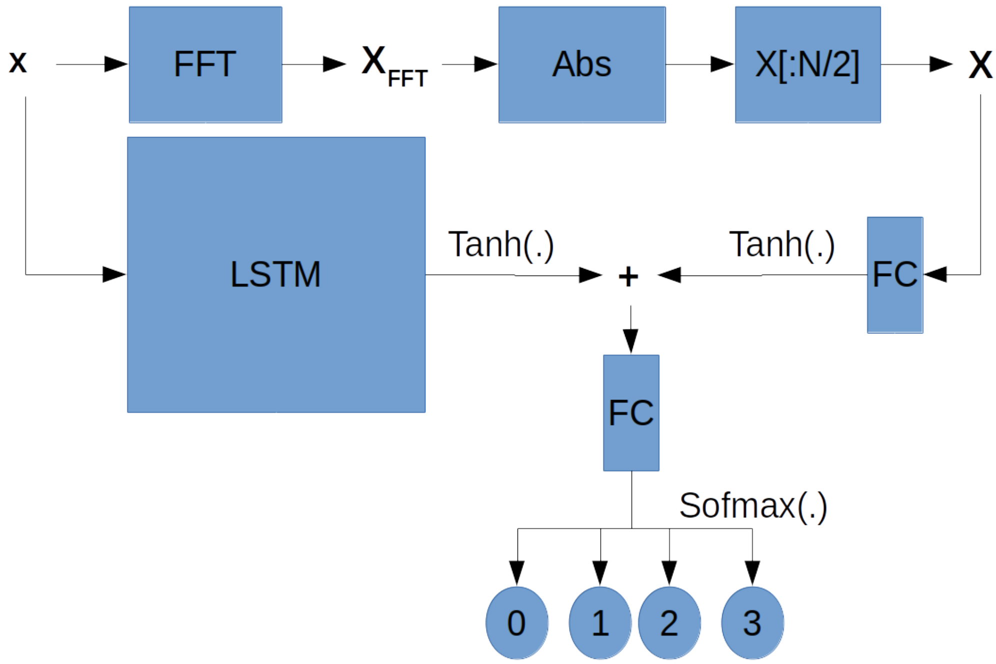

Figure 3: FFT LSTM model used

In addition to the basic LSTM model described in Figure 2, we also produced a more advanced model which played on the idea that there is added information on the frequency domain [2], [3]. This model is described in Figure 3.

The general idea is to compute the Fast Fourier Transform of each ECG input strip, take the absolute value, and one size of the transform (meaning the information from $[0, \pi]$). This idea stems from the fact that cardiac signals are periodic in nature, and display peaks in particular frequencies, as noted in Figure 4.

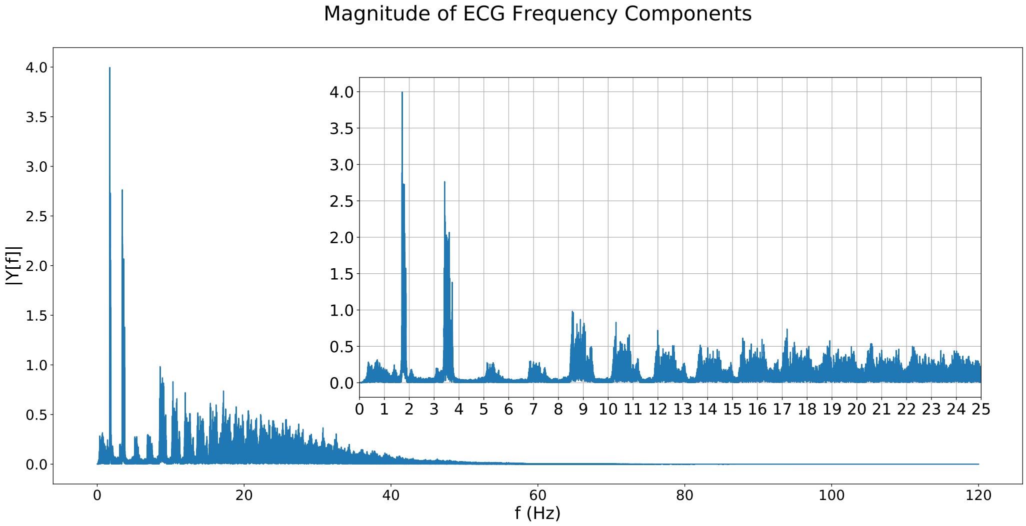

Figure 4: Demonstration of the frequency domain of an ECG signal

After computing the FFT, we feed the resulting data to a fully connected layer. Later, we apply a $\tanh(\cdot)$ to this result. After this point, we merge this result with the LSTM signal flow described in the basic model. It becomes a sequential time result regularized by the frequency information. This then is moved through a fully connected layer, and finally a softmax.

3 Experiment and Results

In this section we will expose more details about the experiment and explore the results. The section will be divided as follows: first we will mention general experiment procedures that apply to both cases. Then we will expose results that pertain to the specific models achieved.

The dataset containing the AFIB data on its own was not enough to provide enough data for the models to be able to learn consistently. This was partly due to its highly imbalanced nature. The original class distribution was as follows:

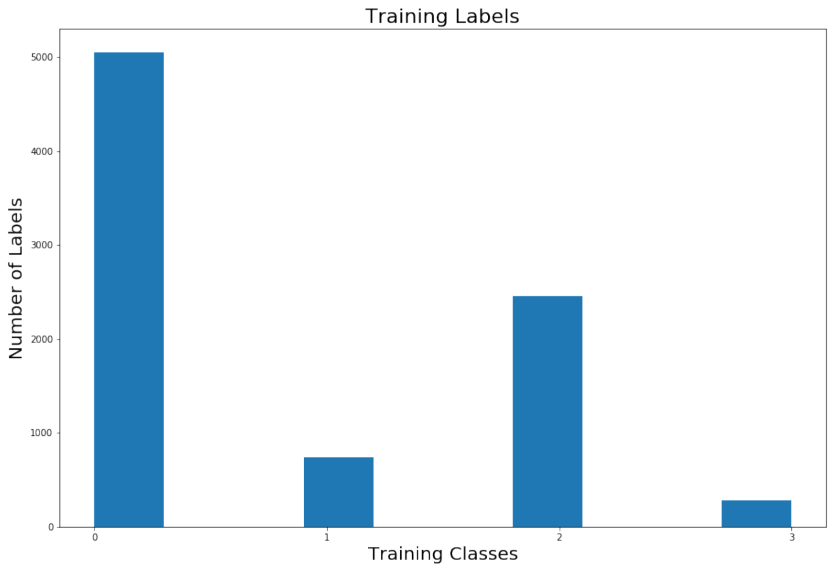

Figure 5: The original class distribution contained mostly normal rhythm waveforms, with few examples of AF and noisy ECG strips.

Therefore, in order to mitigate this situation described Figure 5, and help the model learn, the following augmentation scheme was proposed: double the number of classes which are not as well represented with examples that have a small amount of random noise. This is the equivalent to collecting a similar ECG strip, however, now with some degree of noise.

In the case of the noisy waveforms, we can allow more slack on the amount of noise. Therefore we produced two batches of normal noise defined as follows: $\mathcal{N}(0,1)$, where the resulting augmented dataset was:

$$\mathbf{X}_{aug,3} = \mathbf{X} + \mathcal{N}(0,1)$$

In the case of the AF waveforms, we have to be careful not to incur too much noise. Therefore, the resulting augmentation set was as follows:

$$\mathbf{X}_{aug,1} = \mathbf{X} + \mathcal{N}(0,0.3)$$

Finally, the augmented dataset was concatenated with the original dataset. The resulting class distribution follows:

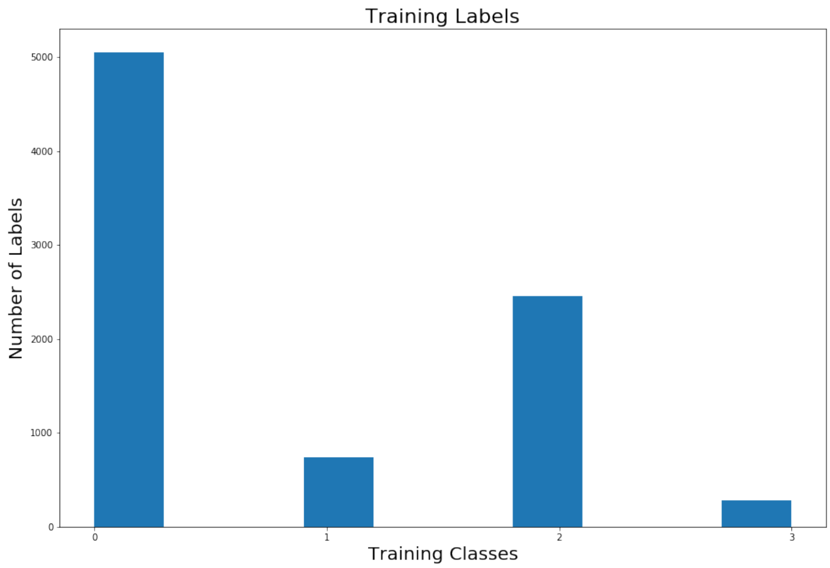

Figure 6: The augmented class distribution shows a four times more AF examples, as well as four times more noisy examples.

With this data augmentation scheme, it was possible to achieve significantly more learning in both models.

3.1 Basic LSTM Model Results

The basic LSTM model ended up producing the best results. The training accuracy reached 95% accuracy, while the test set accuracy reached an impressive 92% accuracy.

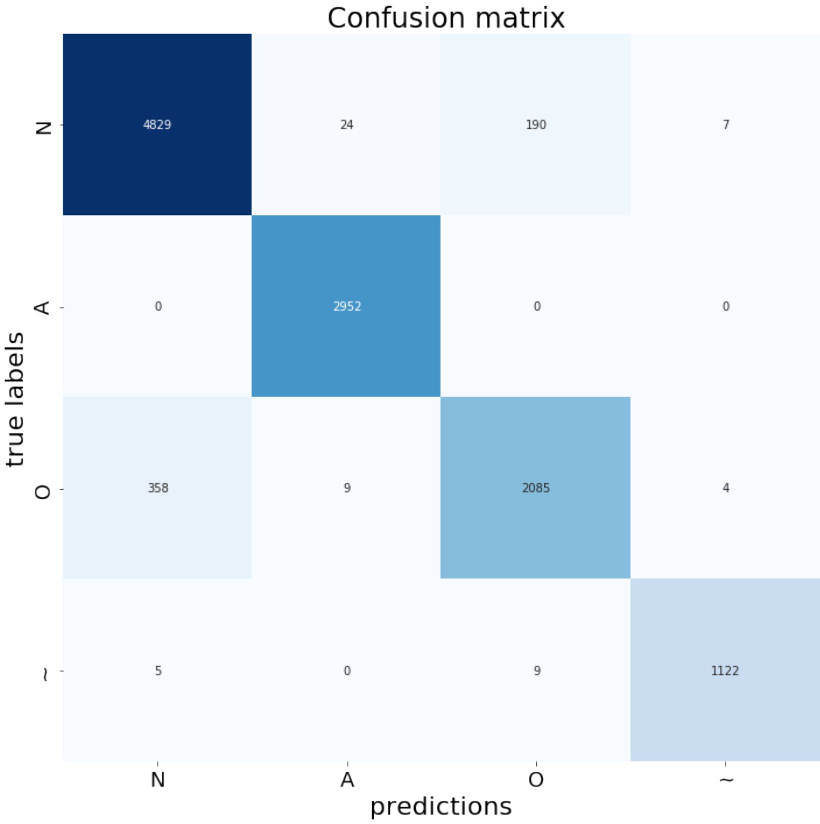

Figure 7: Confusion matrix produced by the basic LSTM model

Moreover, the table containing the different metrics for training:

Class label	Precision	Recall	F1-Score	Support
0	0.93	0.96	0.94	5050
1	0.99	1.00	0.99	2952
2	0.91	0.85	0.88	2456
3	0.99	0.99	0.99	1136

Table 2: Metrics pertaining to the training results.

The results presented in Figure 7 and Table 2 show that the model was able to appropriately learn all classes during training. Next, the test results:

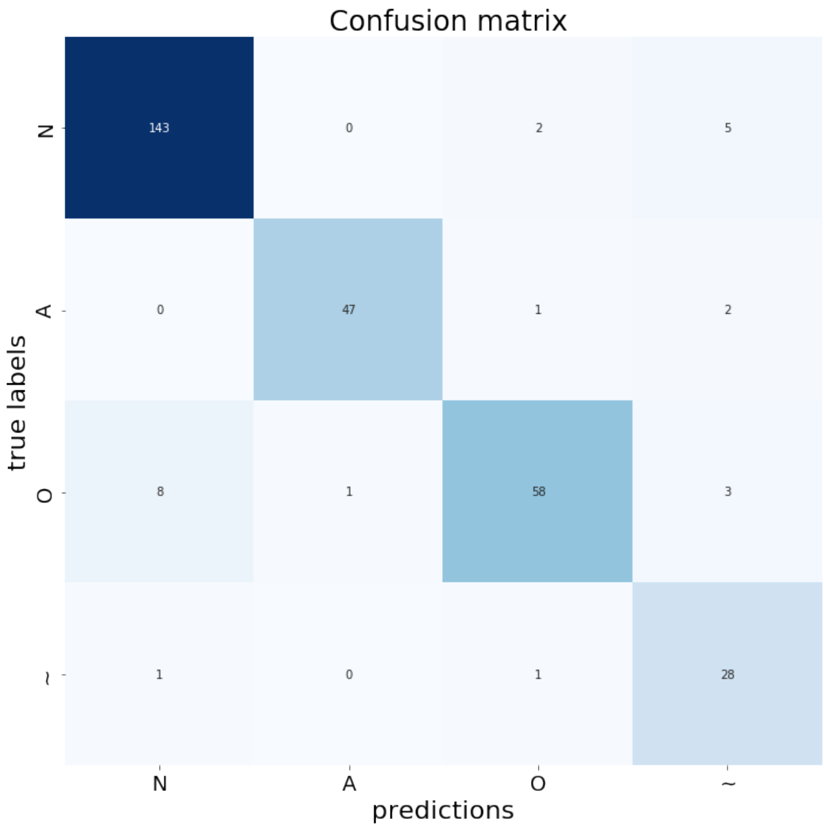

Figure 8: Confusion matrix produced by the basic LSTM model test dataset

Moreover, the table containing the different metrics for test:

Class label	Precision	Recall	F1-Score	Support
0	0.94	0.95	0.95	150
1	0.98	0.94	0.96	50
2	0.94	0.83	0.88	70
3	0.74	0.93	0.82	30

Table 3: Metrics pertaining to the test results.

As it was mentioned previously, the basic LSTM model produced a considerably good performance in training and validation. We estimate that if we increase some of the classes so that they are completely even, performance would increase.

3.2 LSTM Model with FFT Results

Now, the LSTM model with FFT produce respectable results. However, it did not perform as expected. The initial expectation was for this model to outperform the basic model, however, this was not the case. After some analysis, the consensus was that the augmentation process needs to be re-designed, so it may better suit this sort of model. Nevertheless, what follows is the training results:

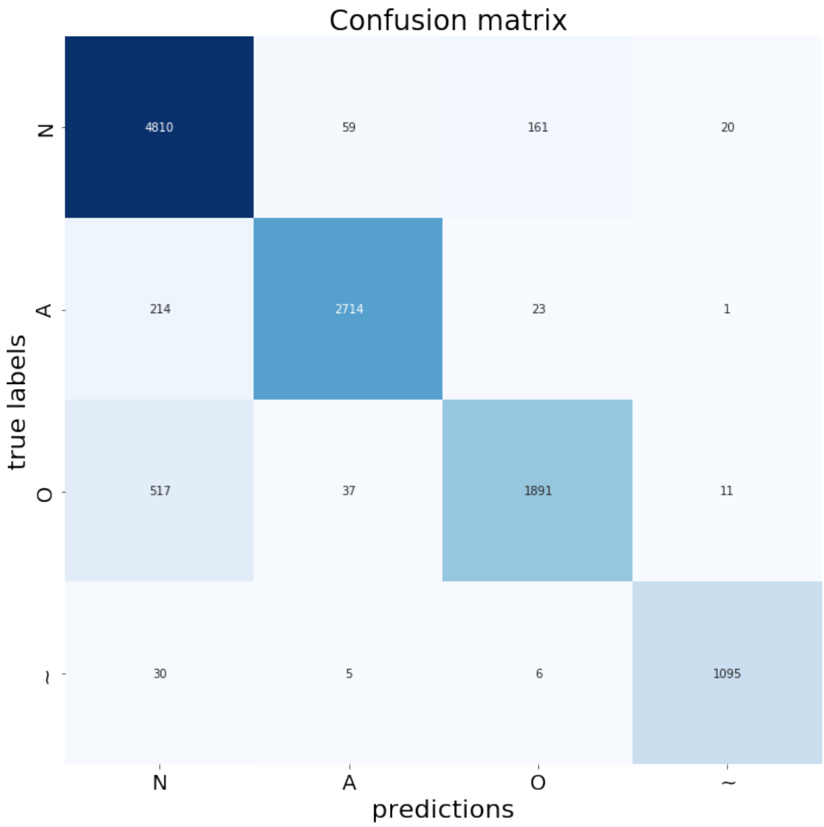

Figure 9: Confusion matrix produced by the basic LSTM model

Moreover, the table containing the different metrics for training:

Class label	Precision	Recall	F1-Score	Support
0	0.86	0.96	0.91	5050
1	0.96	0.92	0.94	2952
2	0.91	0.77	0.83	2456
3	0.97	0.96	0.97	1136

Table 4: Metrics pertaining to the training results.

The achieved training accuracy was 91%. This demonstrates that the model was able to learn the training data. The confusion matrix in Figure 9 demonstrates the misconceptions of the model, more interestingly between normal waveforms and other sinus rhythms. Moreover, Table 4 shows the metrics, confirming some of the confusion in class 2 (other sinus rhythms).

Now, the test results:

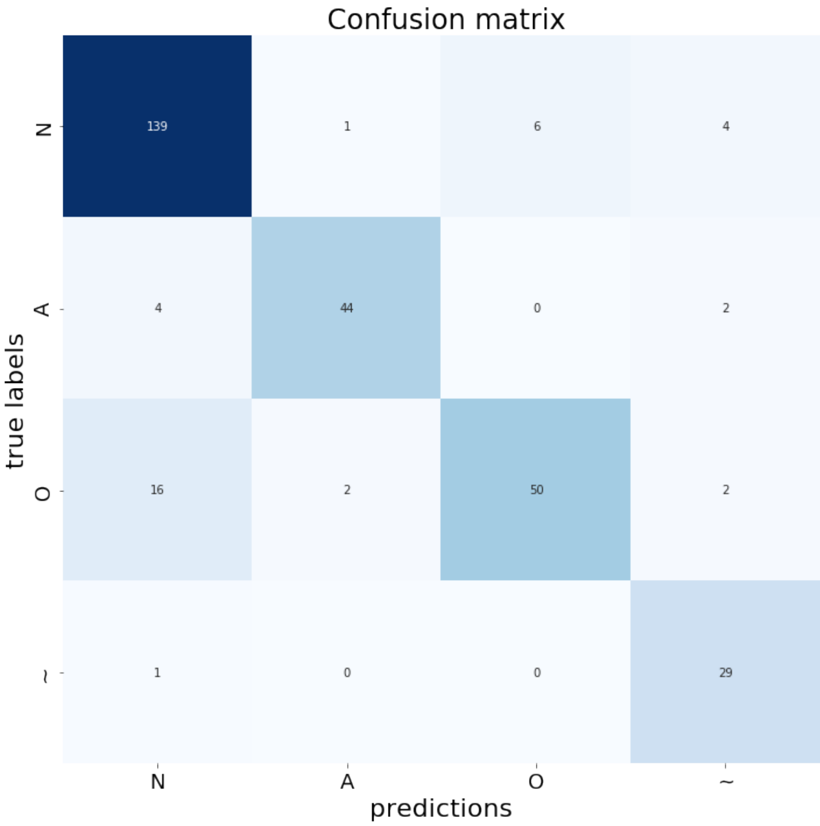

Figure 10: Confusion matrix produced by the FFT LSTM model on the test dataset

Moreover, the table containing the different metrics for test:

Class label	Precision	Recall	F1-Score	Support
0	0.84	0.93	0.90	150
1	0.94	0.88	0.91	50
2	0.89	0.71	0.79	70
3	0.78	0.97	0.87	30

Table 5: Metrics pertaining to the test results.

The test set performance showed in Figure 10 and Table 5 demonstrate that the model was still able to learn. However, performance was not as good as in the basic LSTM model.

4 Conclusion

This project allowed the opportunity to work with a very interesting dataset, whose purpose is noble. The idea to try to create a more innovative solution that will allow people a higher chance of survival is very stimulating. By proposing two different models, whose test performance are respectable, we were able to learn about Atrial Fibrillation and contribute to a healthcare problem. While the performance of the LSTM FFT model was not in par with the expected, we estimate that by putting more time and work into the augmentation scheme, we should be able to produce better results. In any case, the simplicity and considerable performance of the basic model gives assurance that we are at least providing an alternative to save lives.

References

[1] G. Clifford, C. Liu, B. Moody, L.-w. Lehman, I. Silva, Q. Li, A. Johnson, and R. Mark, "Af classification from a short single lead ecg recording: the physionet computing in cardiology challenge 2017,"

[2] L. Sörnmo and P. Laguna, *Bioelectrical signal processing in cardiac and neurological applications*, vol. 8. Academic Press, 2005.

[3] R. W. Schafer and A. V. Oppenheim, *Discrete-time signal processing.* Prentice Hall Englewood Cliffs, NJ, 1989.