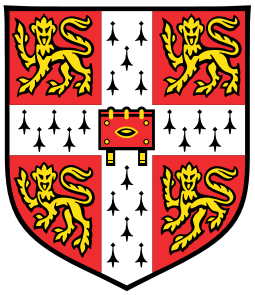
Data of Your Heart: Screening for Atrial Fibrillation



MEng Project Report

Jordan Smith
JS2432

Supervisor: Dr Elena Punskaya

Division F Engineering Department
University of Cambridge, 2022
United Kingdom

June 2022

Disclaimers

Student Disclaimer

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Author: Jordan Elliot Smith
Signed:

Abstract

An abstract.

Contents

1 Introduction

1.1 The Problem

Atrial Fibrillation (AF) is a common abnormal heart rhythm that is associated with a five-fold increase in stroke risk [2]. After being recommended to hopsital for an ElectroCardiogram (ECG) scan, a patient can be diagnosed with AF if the signifying features of AF are detected in the ECG taken during monitoring. If a patient is diagnosed with AF, medication can be administered to reduce the stroke risk. Currently, however, patients only have their heart monitored if they are symptomatic with AF, such as experiencing heart palpitations, yet a significant proportion of AF patients will be asymptomatic, therefore not referred for examination.

Furthermore, some patients will only experience AF episodes, which are short lasted episodes of AF which occur at varying frequencies, rather than consistently show signs of AF. These episodes are often not detected during ECG monitoring in hospital, and therefore the diagnosis is missed despite the risks of AF still being present.

AF is identified through irregular-irregular RR intervals and an absence of P waves in an Electrocardiogram (ECG), see Figure 1. This ECG is taken at a single time point during a hospital visit, therefore it is likely that an AF episode is missed, especially for patients with low AF burden.

Figure 1: Electrocardiogram (ECG) recording with Zenicor device [3], showing Atrial Fibrillation (AF) (a) and sinus rhythm (b).

Screening offers a method to identify these asymptomatic, low AF burden patients who are otherwise neglected by current methods, but still at risk. AF burden is a measure of the proportion of time a patient experiences AF episodes, and it increases as the age of the patient increases [4], meaning the risk of AF continuously increases and should be monitored. The SAFER study [5] is looking into using the Zenicor 1 lead ECG device for patients to use at home, unassisted, to take 30-second samples 4 times a day over 3 weeks, at low cost. This significantly increases the chances of detecting AF episodes in low AF burden patients, and therefore enables the correct treatment or monitoring of these at risk patients.

However, this generates a significant number of samples, with 84 per patient, which need to be checked by a Cardiologist. Currently, the Cardiologist searches through all of these samples and if at least one sample out of 84 shows AF signs, the diagnosis is made. With only one AF sample needing to be found, this process can be considerably more efficient if this AF sample is the first one seen by the cardiologist. This is the motivation for this project.

1.2 The Challenge

While increasing the time frame of the ECG, screening using a 1 Lead ECG has issues of increased noise and decreased coverage of the heart's electrical activity.

The increased noise arises from multiple environmental factors, including user error. A 12 lead ECG taken in hospital is generated with trained professionals using sophisticated equipment, whereas a 1 lead ECG will be self administered. The resulting generated samples will be susceptible to contaminated contact points, movement of the patient, possible external signals such as from nearby electronic devices. This leads to a signal which is noisy, has baseline wander, and will have high frequency components not seen in 12 lead ECG.

Furthermore, the decreased coverage of the hearts electrical activity is due to the 12 Lead ECG taking signals from multiple angles across and through the heart, as seen in Figure 2, whereas a 1 lead ECG only detects transverse signals through the heart, labelled as "1" in Figure 2. This leads to significant losses in the signals that can be detected. In particular, "p" waves which would be found in a 12 lead ECG may be completely undetectable in a 1 lead ECG, which could incorrectly be used to diagnose AF when Atrial Depolarisation, the process which causes the p waves, is actually occurring.

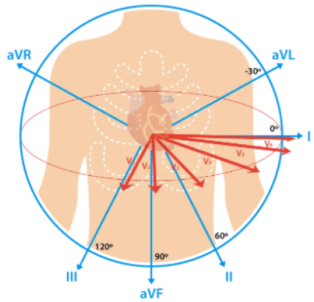


Figure 2: 12 lead ECG signal locations

Another consideration for this project is the presence of other heart arrhythmia in the samples, such as Bradycardia, Tachycardia and Heart Block. These can easily be confused with AF, and therefore the distinction between them is important.

1.3 The objective

To develop a systematic way to order the samples by likelihood of exhibiting Atrial Fibrillation symptoms, in order to speed up the review process for the Cardiologists in the SAFER trial.

Prioritisation of the samples most likely to exhibit AF signs is the aim of the project.

2 Literature Review

2.1 Background

A highly active area of research, classifying ECGs for AF or Sinus (normal) Rhythm (SR) has seen numerous academic competitions, with the largest scale being the Physionet Challenges. Through these competitions, many approaches to this classification have been developed and implemented for use on 1 lead ECG data. Traditional approaches involve more hand crafted feature detection, with approaches using Signals Processing techniques to detect the different features of the ECG signal and then running deterministic algorithms on these feature properties.

2.2 Traditional Approaches

One particular, more traditional approach, is the one proposed by Dr Jie Lian et al. in their paper "A Simple Method to Detect Atrial Fibrillation using RR Intervals" [6], classifies ECGs using only the RR intervals of the ECG. It takes the RR intervals, calculates the difference in RR intervals at each step, and plots these two properties with a grid segmentation of the values to count the proportion of grids which are populated by at least 1 point in a scan of given time length. This method was tested on the MIT-BIH Atrial Fibrillation database [7] which is comprised of data gathered using a Holter device. A Holter device produces significantly different ECG samples to a 1 Lead handheld ECG device. The presence of 'extra peaks' which would be identified in a noisy samples could reduce the accuracy of this method.

2.3 Modern Machine Learning techniques

In the past five years more focus has been placed on Machine Learning (ML) techniques, and especially Deep Learning (DL), to be used for the task of ECG classification. The ability of these methods to function well on noisy ECGs has been especially beneficial. The DL methods increasingly display the capability to learn the waveform shapes, both small and large scale features that correspond to AF or SR, and in some proposals the ability to distinguish between AF and other heart arrhythmia, along with identifying samples that are too noisy. This is especially useful for the application to the SAFER study, where the expectation is that plenty of noisy samples will be produced by the Zenicor device due to incorrect operation, and the presence of other heart arrhythmia is likely, therefore a model which distinguishes between these classifications is very useful.

In 2017, one of the aforementioned Physionet challenges took aim at classification of noisier Lead 1 ECG samples, with data collected from small handheld devices, only differing from the SAFER trial data in that they are variable length, with some less than 10 seconds and other longer than a minute. The challenge asked for methods to classify samples between "Normal" "Atrial Fibrillation" "Other heart arrhythmia" and "Noisy" categories, measuring the performance as the average F1 score over each category. Two of the winning submissions to this challenge were investigated in depth.

1. "ENCASE: an ENsemble ClASsifiEr for ECG Classification Using Expert Features and Deep Neural Networks" [8]. This method combined the use of traditional feature detection from Statistics, Signals Processing and Medicine, with modern DL methods that learn features through data-centric approach. This proposal method had a high accuracy with an F1 score of 0.83. This method was not pursued for the project, however, due to most of the methodology being omitted from the report submission. Therefore applying this method from scratch, without assistance from trained Cardiologists, is beyond the scope of this project.

2. "Robust ECG Signal Classification for Detection of Atrial Fibrillation Using a Novel Neural Network" [9], applied a more DL centred approach, with a novel form of the famous ResNet [10] neural network, using residual connections, being applied for the ECG classification. This method is particularly interesting because not only did it generate exceptional results in the challenge, with a near winning F1 score of 0.82, but it also had a complete architecture and algorithm explanation included in the report.

3. The paper, Identification of patients with atrial fibrillation: A big data exploratory analysis of the UK Biobank [11], analysed the performance of 10 ML techniques with some being classical ML approaches using Support Vector Machines, and others being a combination of classical ML with DL approaches. On the subset of the UK Biobank dataset the combination approach proved to be the most effective at ECG classification. This approach, however, also included the used of expert features assumed beyond the scope of this project.

3 Methodology

3.1 The Data

The primary data source for this project is the dataset produced in the SAFER study [5], which contains over 9988 samples that have been individually reviewed by cardiologists, and growing as the study continues. As well as this, these datasets contain around 175000 samples in which the patient themselves has been diagnosed as having AF or not. In these samples, there will be numerous that are ground truth labelled as 'AF', yet do not actually display any signs of AF, because those individual samples are from a patient with low AF burden and therefore not all the samples from this patient will show AF signs.

Another source of data is 2 open-source datasets, the Physionet Challenge 2017 dataset (8528 samples) which is Lead 1 ECG data, and the China Physiological Signal Challenge (CPSC) Database [1] (6877 samples) which is 12 lead ECG data. The 12 lead data is not as useful, even though the lead 1 data can be isolated, because it is taken from ECGs produced in a hospital, therefore, is less noisy. Augmentation to add noise, and bandpass filtering, in order to simulate data produced by a handheld device is needed.

Furthermore, another consideration of the datasets is the skew towards 'Non-AF' samples. For the SAFER dataset, there is an average of 12%, of samples labelled, labelled as 'AF'. For the Physionet 2017 dataset only 9%. Care is taken to prevent the model from overfitting to the 'Non-AF' end of the probability spectrum.

Finally, the presence of samples showing the aforementioned other heart arrhythmias is another consideration in each dataset. These can be easily confused with AF if the model is incorrectly trained and does not recognise the correct features. With these arrhythmias often present in the general population, it is very important for the model to be able to distinguish between them, and as such the careful consideration of how to train the model and which classifications it should output will be needed.

3.2 Novel Neural Network approach

The Novel Neural Network (NNN) approach [9] chosen has architecture as shown in Figure 3. This network was applied to samples split into 5-second sections, with each section having a classification prediction made for it, and then the classification that appeared the most over all of these split samples is the one chosen as the classification for the entire sample. Although this method was initially applied to a classification problem, the model will be adapted to produce a probability, rather than classification, of AF. There are multiple options for achieving this, with the simplest taking the proportion of the sample splits which are classified as AF as the likelihood of the sample showing AF. A more likely option to achieve better, more continuous, results would be to examine the output of the softmax layer, and through experimentation find a suitable way of combining these outputs into a confidence of AF.

The original NNN model produced for the Physionet challenge 2017 was programmed to classify between 'AF', 'N' (Sinus rhythm), 'O' (Other heart arrhythmias) or ' ' (Noisy sample). The work of this project is only interested in 'AF' or 'Not AF', therefore this has been adapted accordingly. Currently 'O' and 'N' are combined, and '∼' are not of interest

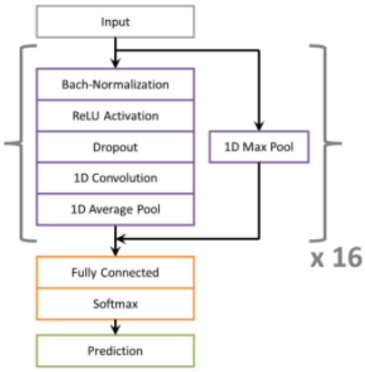


Figure 3: Proposed network architecture for the Novel Neural Network approach with skip connections [9]

other than to state that the model was not confident for samples labelled '∼'.

Classification	F1 Score	Support
A	0.88	743
N	0.93	5070
O	0.84	2464
∼	0.72	286
Accuracy avg	0.84	
Weighted avg	0.89	

Table 1: Results of NNN model from Physionet challenge 2017 tested on its train dataset for quick validation

Classification	F1 Score	Support
A	0.98	918
N	0.90	1221
Accuracy avg	0.94	
Weighted avg	0.93	

Table 2: Results of NNN model from Physionet challenge 2017 tested on lead 1 CPSC data [1]

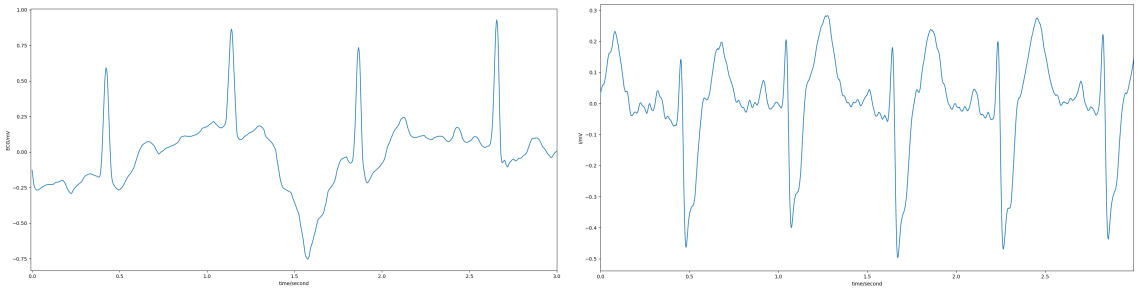
4 Results

4.1 Tested on Physionet Challenge 2017 dataset

In order to check that the NNN model worked it was tested on the dataset on which it was trained, see Table 1 for results. Results were promising, with values confirming the model functioned as expected. The mean F1 score of 84.25 is slightly higher than the competition score of 82, which is to be expected because the model had already seen this data, and so higher accuracy is expected.

4.2 Tested on CPSC database [1] Lead 1 data

The next step was to apply this model to data that it had not seen before, the CPSC database. Testing on this data, which was collected in a hospital by trained professionals, is not fully representative of the model performance for the application to data collected using the Zenicor device for the SAFER trial but is still a useful tool for validation purposes, see Figure 4 for comparison. Results are shown in Table 2, with only the classifications of 'A' or 'N' of interest. This model actually performs significantly better on this dataset than when tested on its own train dataset, which is initially surprising. This is explained by the lower noise present in the samples from the CPSC database, due to the higher quality equipment, leading to the model performing very well at recognising key features, because they are more clear.



(a) Physionet challenge 2017 sample

(b) CPSC sample

Figure 4: Samples from the two open-source datasets, showing the benefit of the ECG taken at hospital with more clearly defined p, qrs, and t waves (4b), and the drift and noise sometimes found in samples taken from self administered 1 lead ECGs (4a).

5 Discussion

A discussion.

6 Conclusion

A conclusion.

References

[1] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He, J. Li, and E. N. Y. Kwee, "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," *Journal of Medical Imaging and Health Informatics*, vol. 8, pp. 1368–1373, 8 2018.

[2] NHS, "Complications - atrial fibrillation," 5 2021.

[3] S. Schukraft, M. Mancinetti, D. Hayoz, Y. Faucherre, S. Cook, D. Arroyo, and S. Puricel, "Handheld ecg tracking of in-hospital atrial fibrillation the hecto-af trial clinical study protocol," *Trials*, vol. 20, p. 92, 12 2019.

[4] I. C. V. Gelder, J. S. Healey, H. J. Crijns, J. Wang, S. H. Hohnloser, M. R. Gold, A. Capucci, C.-P. Lau, C. A. Morillo, A. H. Hobbelt, M. Rienstra, and S. J. Connolly, "Duration of device-detected subclinical atrial fibrillation and occurrence of stroke in assert," *European Heart Journal*, vol. 38, pp. 1339–1344, 5 2017.

[5] SAFER, "The safer trial screening for atrial fibrillation with ecg to reduce stroke."

[6] J. Lian, L. Wang, and D. Muessig, "A simple method to detect atrial fibrillation using rr intervals," *American Journal of Cardiology*, vol. 107, pp. 1494–1497, 5 2011.

[7] G. Moody and R. Mark, "Mit-bih atrial fibrillation database," 11 2000.

[8] S. Hong, M. Wu, Y. Zhou, Q. Wang, J. Shang, H. Li, and J. Xie, "Encase: An ensemble classifier for ecg classification using expert features and deep neural networks," vol. 44, pp. 1–4, IEEE Computer Society, 2017.

[9] Z. Xiong, M. K. Stiles, and J. Zhao, "Robust ecg signal classification for detection of atrial fibrillation using a novel neural network," vol. 44, pp. 1–4, IEEE Computer Society, 2017.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 12 2015.

[11] J. Oster, J. C. Hopewell, K. Ziberna, R. Wijesurendra, C. F. Camm, B. Casadei, and L. Tarassenko, "Identification of patients with atrial fibrillation: A big data exploratory analysis of the uk biobank," *Physiological Measurement*, vol. 41, 2020.

Appendix

Appendix goes here.