Department Of Mechanical Engineering

MECH0020 Individual Project

AY 2022/23

Word Count: 7498 words

Student: (Name and ID)	**Emanuel Cristian Ciudin (Student ID: 20001230)**
Project Title:	**A Deep Learning ECG Algorithm for Prediction of Atrial Fibrillation and Major Adverse Cardiovascular Events in Individuals without Prevalent Cardiovascular Disease**
Supervisor:	**Dr. Ryo Torii** **Dr. Michele Orini (External Supervisor)**

Declaration

I, Emanuel Cristian Ciudin, confirm that the work presented in this report is my own. Where information has been derived from other sources, I confirm that this has been indicated in the report.

Abstract

The electrocardiogram (ECG) records the electrical activity of the heart which is undergoing a continuous depolarization-repolarization cycle. Medical professionals currently use the ECG to assess and diagnose certain cardiac conditions, such as arrhythmias and ischemic events.

Cardiovascular disease is the leading cause of death of worldwide and the growing sedentary lifestyle will only further deteriorate the situation. Cardiovascular health is now more relevant than ever and early detection of disease signs would improve the quality of life for millions of individuals.

Although, doctors show a high accuracy on diagnostic prevalent conditions, especially for common affections such as atrial fibrillation, prediction of incident diseases is a challenging task for the human eye. Risk factor models have been developed to track incident cardiovascular disease types, but they rely on human-made features and so may prove to be sub-optimal.

Machine learning and deep learning models have made tremendous progress in the last decade, with many medical applications currently in development. Convolutional neural networks, a deep learning model, have shown great applicability to dynamic images, such as the ECG, and have proven competent on detecting prevalent cardiovascular diseases with a high accuracy.

The challenge now lies into creating deep learning models for the prediction of prevalent diseases. Current research shows progress in developing algorithms on 1-lead and 12-lead ECGs for both detection and prediction, but problems persist on the architecture of the network and the dataset size.

This project is focused on developing a convolutional neural network architecture based on data provided by the UK Biobank. The project's main objective is to analyse if training a neural network on exercise stress test (stationary bicycle test) data enhances the prognostic value of at-rest ECGs. The targeted incident affections in this study are atrial fibrillation and major adverse cardiovascular events.

Acknowledgements

I would like to deeply thank my two supervisors, Dr. Torii and Dr. Orini, for providing unlimited guidance and feedback throughout the project. Dr. Torii has been my personal tutor since my first year of university and has been instrumental in both my academic and personal success and development in these past 3 years. Dr. Orini (UCL and Barts Heart Centre) offered continuous support throughout the project and his expertise and insights were essential in writing a high-quality paper.

Secondly, I would like to offer my appreciation to the half a million participants in the UK Biobank study, from which I procured data for the project. Their voluntary involvement in the medical study provided incredibly valuable research data to which I was thankful to have access to.

This project is dedicated to my late father who passed away last year. He was a cardiologist specialised in electrophysiology and I would hope he's able to see from up there the advances I made with this project is his field of work. He was a great man, father, and an incredibly kind individual who guided me throughout my life and led me to this moment.

Lastly, I would like to thank my whole family for their constant support and love and my roommate, Raphaël and my UCL friends. My mom, who is an incredible 'warrior', has always fought so I could have the best upbringing and for that I am eternally grateful.

Nomenclature

AF	Atrial Fibrillation
MACE	Major Adverse Cardiovascular Events
HAF	Combined Dataset of Healthy (H) and AF
HMACE	Combined Dataset of Healthy (H) and MACE
1-Dimensional	1 column of at-rest ECG data
5-Dimensional	5 columns of multiple exercise stages of the ECG stress test

Contents

List of Tables

List of Figures

Project Background

The cardiovascular system is essential to human beings for efficient transport of nutrients, blood flow and multiple other functions.[1] The heart is at the core of it: a muscle acting as a double pump, with the left part creating the systemic circulation (whole body) and the right part the pulmonary circulation (the lungs).[2] The heart is powered by an electrical conduction system of pacemaker myocardial cells (Figure 1). These have a resting potential and go through a continuous depolarization-repolarization cardiac cycle determined by ionic reactions (Na^+, K^+, Ca^{++}).

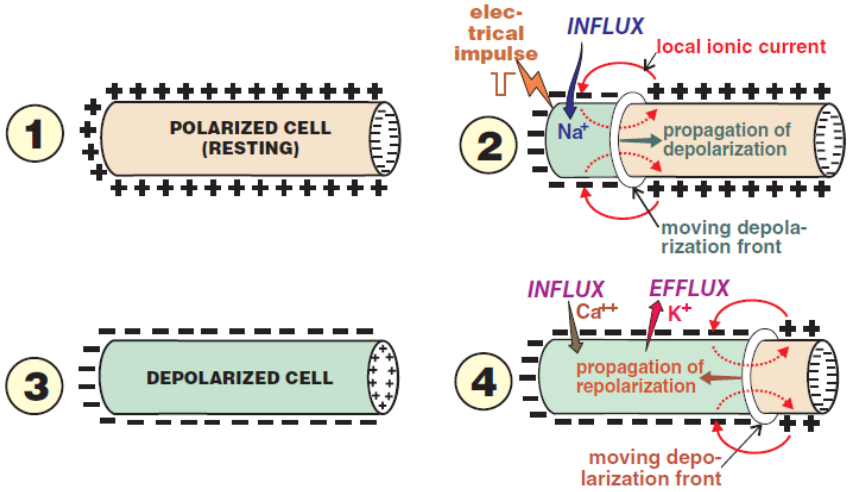

Figure 1: Depolarization and repolarization of pacemaker cells[2]

In a healthy individual, the electrical excitations of the various parts of the heart describe a synchronised pattern. Ischemic cardiovascular events, where sections of the heart don't receive adequate levels of blood and oxygen, are found to directly alter the normal electrical activity of the heart.[3] That is, they create arrhythmias, irregular heart beating patterns, which don't appear in a sinus rhythm (normal heart). For example, atrial fibrillation (AF), a common arrhythmia diagnostic, is found in around 2.1% of the UK population.[4] Some mild symptoms of ischemic conditions are angina pectoris (chest pain) and dizziness. Myocardial infarction (MI), a serious ischemic event, may lead to sudden cardiac death. Consequently, prevention and prediction methods are essential in clinical cardiology.

Expert medical professionals in electrophysiology, make use of the electrocardiogram (ECG) as a diagnostic tool for such conditions and many others. The ECG presents a visual analysis of the electrical waveform signal coming from the heart. Consequently, it records the heart's continuous depolarization and repolarization under different views (standard ECG has 12 leads-Figure 2).

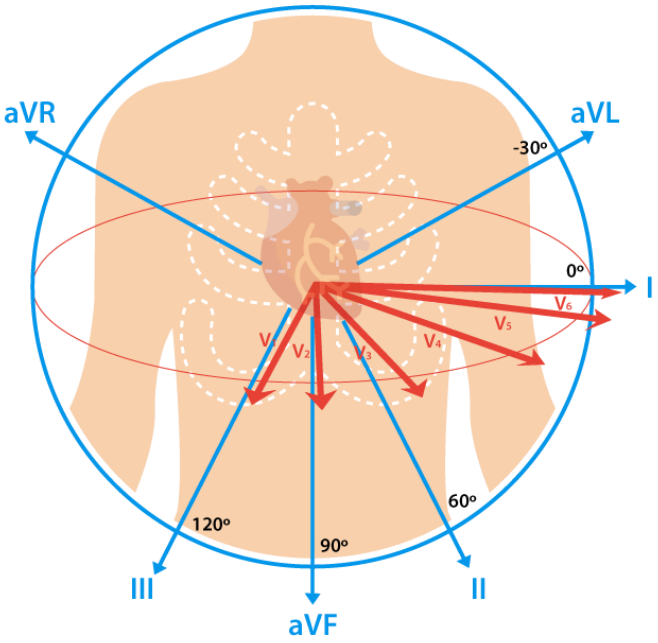

Figure 2: 12 lead views of the heart's electrical activity[5]

General practitioners detected AF with an 80% sensitivity, which might be considered low, but the ECG only forms a part of the investigation conducted on a patient.[6] The use of computer algorithms that can learn complex patterns from data has significant potential to improve many diagnostic procedures such as the ECG. Machine learning (ML) algorithms have proven not only to increase accuracy, but also decrease the diagnostic time. Trained deep neural networks (DNNs), such as one dimensional convolutional neural networks (CNNs), have already been used to identify asymptomatic left ventricular dysfunction (ALVD), something that would require a complex and expensive scan, with 85.7% accuracy based on a 12-lead ECG.[7] AF has also been successfully identified via common machine learning and DNN algorithms (CNN and recurring networks) to provide a comparison of existing methods.[8]

The challenge now persists in predicting future development of cardiac diseases in apparently healthy individuals. Accurate identification of individuals at risk of developing major adverse cardiovascular events (MACE) would allow the implementation of successful prevention strategies in a timely manner. For individuals suffering only from chest pain, without a standing cardiovascular diagnostic, the occurrence of MACE stands at 2.4% and a significant 8.1% for sustained angina pectoris.[9]

Current progress in the field is slowed by a number of reasons. Although machine learning models have been developed to work well on internal datasets, external validation (other hospital/study data) shows suboptimal performance. An optimised input type for maximum information density and an accepted network architecture are also prominent areas of active research.

Project Significance

A prediction algorithm that can classify individuals into groups at risk of developing various cardiovascular diseases could be immensely valuable to healthcare professionals. When a person is diagnosed with AF, the chance of having a stroke is increased 5-fold compared to sinus rhythm.[10]

When a patient presents themselves for a routine check-up with an ECG, the data will be interpreted to assess whether they are at risk to develop a CVD. If an affirmative response is received, they will be closely monitored and prevention measures can be taken, according to professional advice. More frequent follow-up checks will be made to ensure that the risk of serious medical consequences is minimised. CVDs cause 22% of all premature deaths every year in the UK and early detection could potentially save many lives by taking proactive measures.[11]

The proposed diagnosis procedure is presented in the diagram below:

Diagnostic Procedure

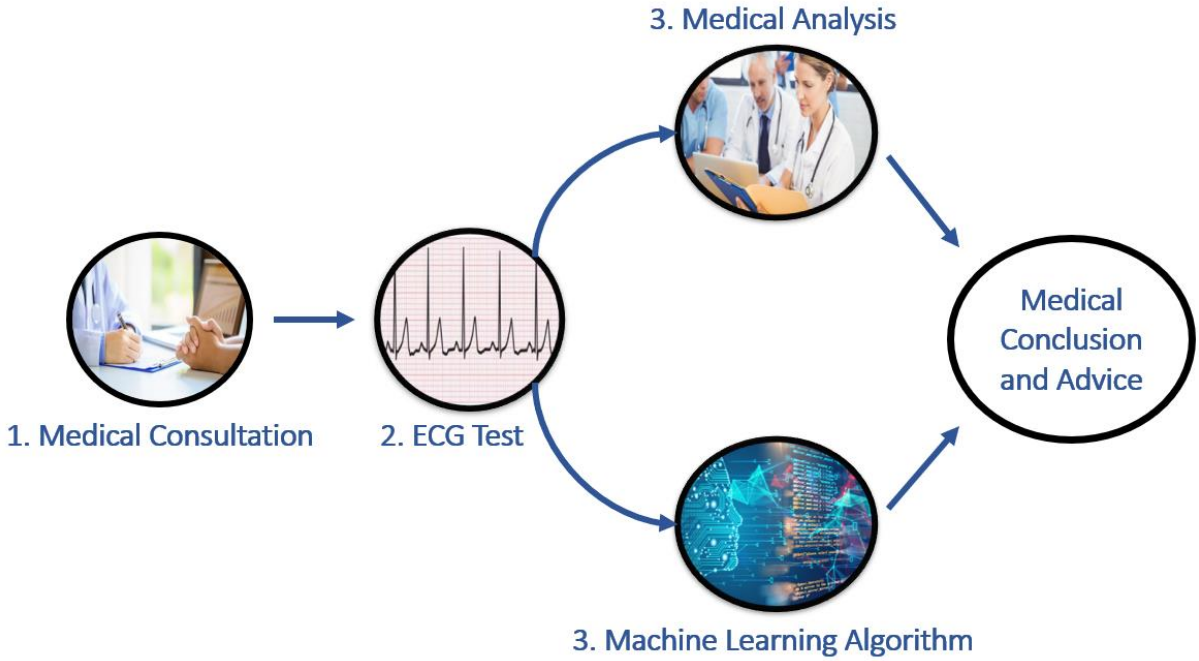

Figure 3: Diagnostic Procedure Flow[12][13][14][15]

Literature Review:

ECG Test Definitions

The ECG is an essential non-invasive medical procedure that assesses an individual's cardiovascular health from the heart's electrical activity. An ECG waveform typically represents the PQRST waves (Figure 4), where each letter corresponds to the electrical change in different parts of the heart. The P wave represents the electrical depolarization of the atria, the QRS complex represents the depolarization of the ventricles and the T wave represents the repolarization of the ventricles.

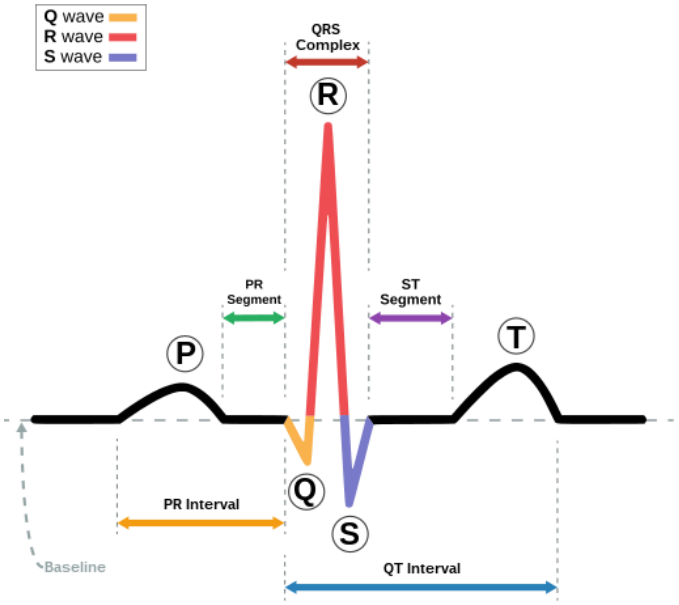

Figure 4: PQRST wave[16]

A standard ECG has 12 leads, 6 are attached to the limbs (limb leads: I, II, III, aVR, aVF, aVL) and 6 chest leads (precordial leads: V1-V6).[2] These provide a full horizontal and vertical plane picture of the heart from which the cardiac rhythm can be identified. Fewer leads, such as the 1-lead and 3-lead ECGs also provide valuable arrhythmias information or are used for long term monitoring purposes. They may be used in tandem with an exercise stress test: a stationary bicycle on which an individual pedals to record the heart's activity under stress. This provides invaluable information for detecting ischemic problems or arrhythmias as they have been heavily associated with an abnormal heart rate recovery post-exercise and induced ST segment depression.[17]

Cardiovascular Electrophysiology Diagnostics and Procedures

Major Adverse Cardiovascular Events (MACE)

The definition of MACE is inconsistent in the medical research literature, therefore this study considers it a composite endpoint of heart failure (HF), myocardial infarction (MI) and life-threatening ventricular tachycardia. This is similar to other literature definitions, with some other considerations being: "Myocardial Infarction (MI), stroke, heart failure (HF) and other ischemic cardiovascular events".[18] It is vastly used in medical studies in randomised controlled trials and observational studies.[19] All the above affections represent can be described as arrhythmias or ischemic events (in

which sections of the heart don't receive adequate levels of blood and oxygen). The two are strongly related, as detailed in the Atrial Fibrillation section.

MACE represents an aggregate of different manifestations of CVD (cardiovascular disease) which are reflected on an individual's ECG. For example, MI can be characterised by an ST-segment elevation, most commonly caused by a blockage of a coronary artery (STEMI).[20] Heart failure also shows different signs on the ECG, depending on the level and location of cardiac disfunction. A common occurrence in heart failure is left ventricular hypertrophy, where the myocardium is enlarged, and is diagnosed in an ECG by an increased amplitude in the QRS complex (best visible in leads V5 and V6).[21]

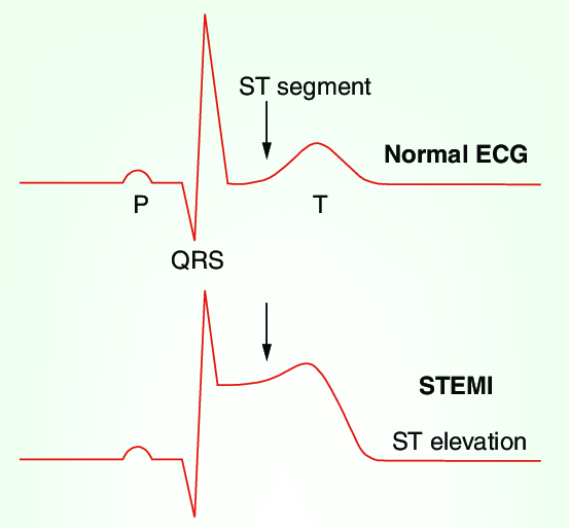

Figure 5: STEMI and normal sinus rhythm with visible ST segment elevation[17]

Atrial Fibrillation (AF)

AF is the most widespread arrhythmia diagnostic, affecting 1.4 million individuals in the UK alone.[4] As per the name, AF is defined by an irregular contraction of the atria, due to electrical impulses received from places other than the SA (sinoatrial) node. Consequently, the atrial disfunction is detectable on an ECG by the absence of the P-wave (atrial depolarization) and an irregular R-R interval. Diagnosing or predicting AF early is crucial as it may be accompanied by various underlying comorbidities. In the Framingham Heart Study which spanned from 1982 to 2012, 57% of the patients who were diagnosed with HF also had prevalent AF.[22] MACE is thus strongly related to AF, with the risk of developing MI being 2 times more when already suffering from AF.[23]

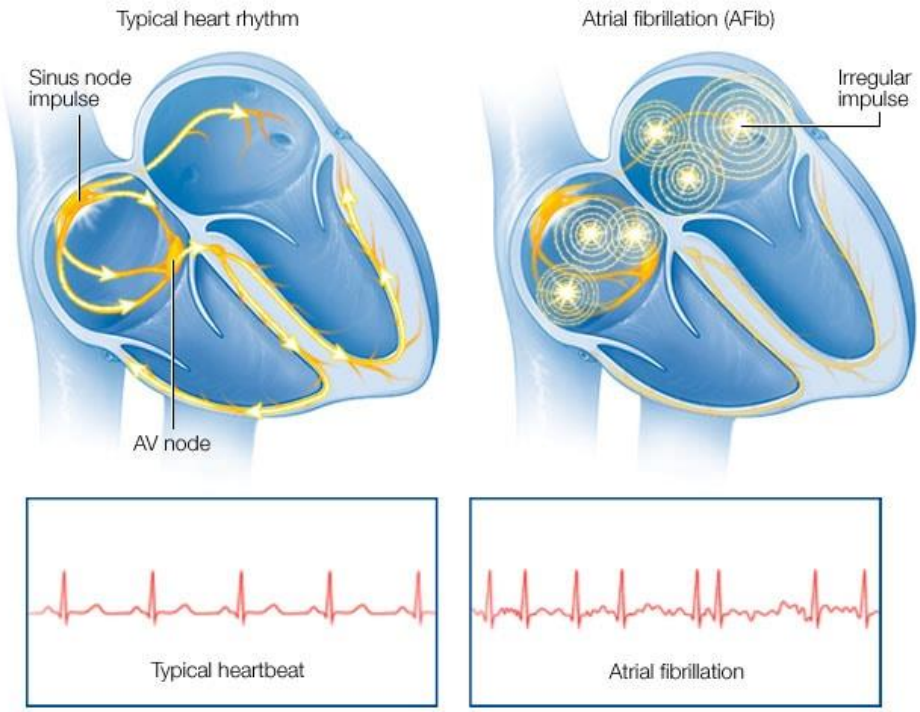

Figure 6: Atrial Fibrillation with irregular R-R, P-wave absence and Normal Sinus rhythm[24]

Statistical Methods and Machine Learning Approaches for the ECG

Statistical Methods

Statistical analysis on ECG has its basis in clinical risk factor models. These models, such as the Framingham Risk score which was derived from the afore mentioned Framingham Heart Study, take into account the following variables: age, sex, body mass, systolic blood pressure, PR interval, HF and hypertension treatment.[25] Using the Framingham Risk score for a 5-year incidence of AF, the C-statistic discrimination varied between 0.66-0.68 depending on the dataset used. The main impracticalities of using pure statistical algorithm to predict incident CVD is that prediction features in the ECG are complex and may vary in symptomatology from person to person. Consequently, human experts may encounter difficulties when deciding on a limited number of features extracted.

Machine Learning and Deep Learning Research

Artificial intelligence applications into the ECG analysis have rose rapidly in recent years, ranging from detection to prediction in 1-lead to 12-lead ECGs. Classical machine learning methods, such as logistic regression, have the same cumbersome problem of manually selecting features seen in statistical algorithms.[26] Deep learning is becoming the new standard, as a result of their efficient automatic feature extraction and their adaptability to deal with raw ECG signals.[27] On detection tasks, CNN performs almost impeccably, with a Tutuko et al. algorithm (AFibNet) reaching 98.97% sensitivity and specificity. As RNN (recurrent neural networks) are more proficient in handling time-series data, as the ECG could be interpreted as, they are also useful detection/prediction tools. An LSTM (Long-Short Term Memory) network solves the vanishing gradient problems of typical RNNs, making it efficient in analysing longer time series such as a higher-sampling frequency ECG.[28] More advanced algorithms are now deployed in research, such as hybrid CNN-LSTM networks. Such an algorithm constructed by Petmezas et al in 2021 achieved 97.87% sensitivity and 99.29% specificity on a multi-class detection problem which included AF.[29]

Research on prediction of incident AF and CVDs on a 5-10 years timeline is still in its early stages. Nevertheless, predictive CNN models have already scored similarly in C-statistics to already established risk factor models.[50]

Current Status and Problems

The biggest impediment faced in developing practical deep learning algorithms is the large amounts of data needed for external validation (validation of trained algorithm on external hospital data). A study which used 221846 12-lead ECGs from 4 institutions reported poor performance on external data on a model which previously yielded great results.[30]

Another main problem is that they require multiple follow-up ECGs, to assess the patient's cardiovascular health. As deep learning algorithms require labelled training data, people must undergo ECG tests at baseline and then they need to be monitored to detect specific outcomes. Only a few studies, such as the UK Biobank, provide access to such valuable training data.

Further investigation is needed to better understand the importance of the input to the algorithm. Although, 12-lead ECGs offer the widest range of information, they may not always be available and a lower number of leads offers diagnostic potential as well. It is also important to note that many CVDs may not show their apparent nature at rest. A fast recovery after exercise was directly associated with lower risk of CVD events.[31] Heart rate recovery association with CVD is still in early research and ECG AI algorithms could provide additional diagnostic support from exercise tests.

Project Aims and Objectives

The **overarching aim** of this project is to build a neural network that is capable of predicating if an individual will develop AF or MACE in a timespan of maximum 11 years. This project also aims to understand if training a neural network on exercise stress test (stationary bicycle test) data enhances the prognostic value of at-rest ECGs. Such an algorithm would be used in parallel with medical professionals to alert them of possible patients that should be more closely monitored in their future check-ups.

The **main objectives** of this project are:

- Pre-processing UK Biobank data to construct useful training, validation and testing datasets.
- Creating statistical data to augment the training dataset.
- Developing a deep learning algorithm in Python using PyTorch to predict the development of AF and MACE in currently healthy individuals with UK Biobank data.
- Testing the validity of the model on discrimination metrics-AUROC (Area under the Receiving Operating Characteristic).

The **workflow** for this study is:

1) Data processing is required to turn the ECG into a useful input for the CNN algorithm. The UK Biobank study presents 66297 raw 1-lead ECGs with a wide range of groups: prevalent AF and MACE, healthy individuals and incident AF and MACE. Filtering through and constructing useful training, validation and testing datasets is required for a working algorithm.
2) Statistical data will be created to provide more training examples and create a balanced dataset for the CNN. This will further increase the accuracy of prediction as the healthy patients outnumber comorbidities roughly 92% to 8%.
3) A binary classification (healthy or suffering from a specific cardiovascular condition) CNN model will be constructed considering reviewed literature and adaptability to the ECG stress test data. This will be developed in Python using the machine learning framework PyTorch and other data analysis libraries such as Pandas and Numpy.
4) Using C-statistic to determine the discrimination of the algorithm via the specificity and sensitivity of the network and its AUROC. This will be compared with current research findings.

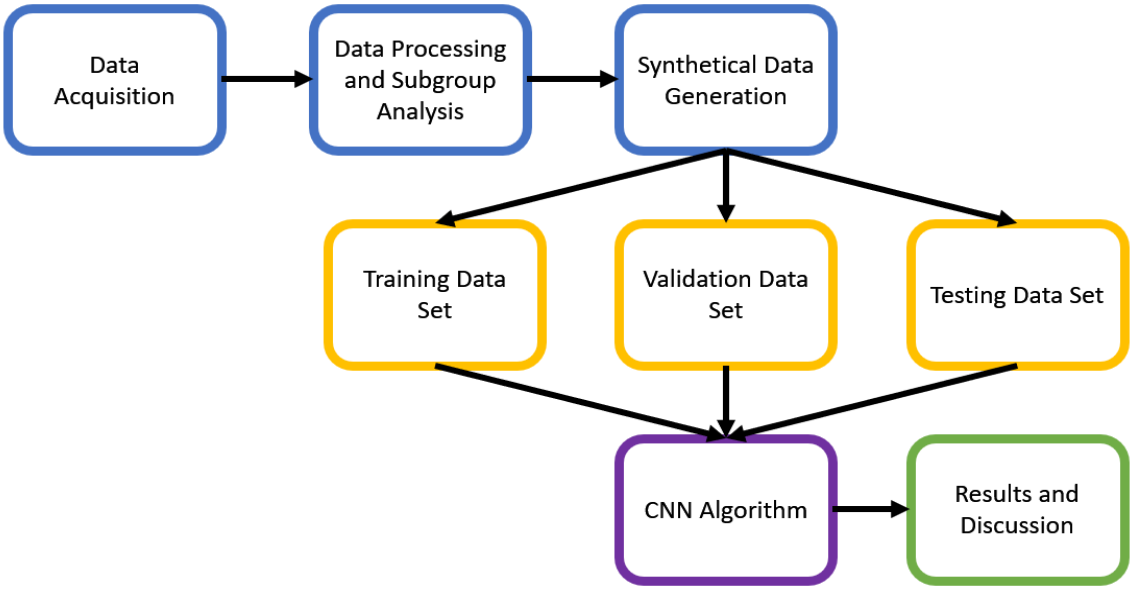

Figure 7: Workflow of the project

Data and Methods

UK Biobank Dataset

The dataset is taken from the UK Biobank Study, a biomedical data base containing health information of half a million participants from the UK.[32] The dataset includes 66297 raw 1-lead ECGs. The ECGs (CAM-USB 6.5, Cardiosoft v6.51.) were recorded on a cycle ergometer with a ramp slope (eBike, Firmware v1.7.).[33] In total 95116 individuals were subjected to the test, with some only completing a part of it considering their Risk Category. For this project only participants with a fully recorded ECG on the exercise stress test were taken into account.

The following format was used to record different parts of the exercise:

Pre-test (at rest)	**15 seconds**
During Exercise (Constant Workload)	2 minutes
During Exercise (Linear Increasing Pedalling Resistance to Peak Power Level)	4 minutes
Recovery (No Pedalling)	1 minute

Table 1: ECG Recording Phases

4 electrodes were placed on the individuals undertaking the test in the following manner:

Electrode I	Right antecubital fossa
Electrode II	Left antecubital fossa
Electrode III	Right Wrist
Electrode IV	Left Wrist

Table 2: Electrode Placement

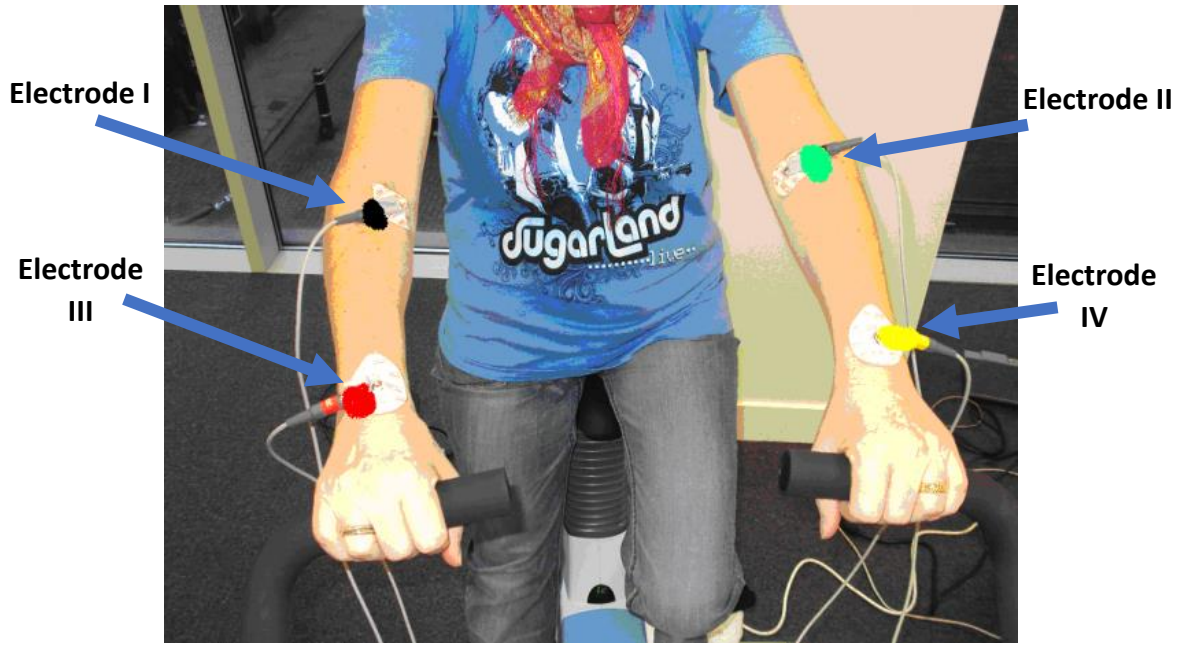

Figure 8: Electrode Placement[34]

Considering the placement of the electrodes to be on the left and right arm, only Standard Lead I is available. The data was received from the ECG device in numerical format, as a function of voltage with

time. The sampling rate of the recording is 500 Hz. The data is completely unfiltered, in pure output form.

Participants were monitored for up to 11 years and health outcomes were obtained from their Hospital Episodes Statistics provided by NHS Digital. The following categories were identified:

Prevalent AF/MACE	The individual suffered from the condition even before recording the test
Incident AF/MACE	The individual wasn't suffering from any conditions initially, but developed one in between 2 recordings
Healthy	The individual is completely healthy in all recordings

Table 3: 3-Classes of Interest

Two sample plots of 4 seconds (2000 time-stamps) recording at rest are showcased in Figures 9,10.

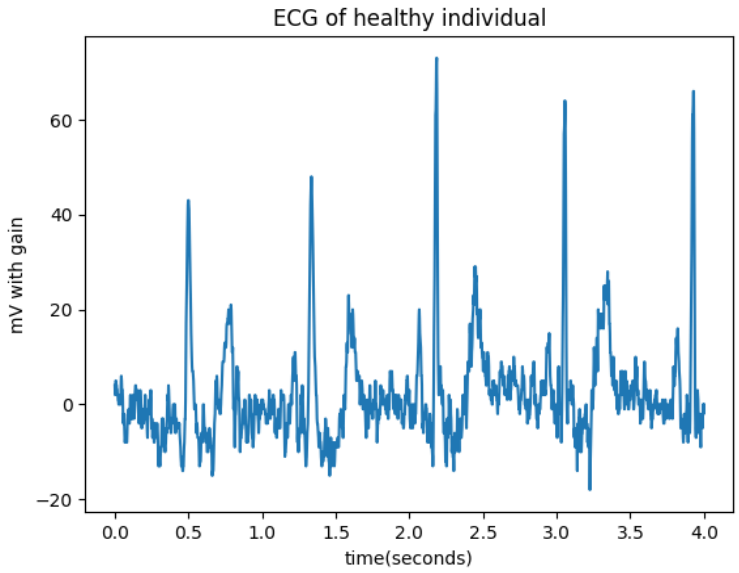

Figure 9: ECG of a healthy individual

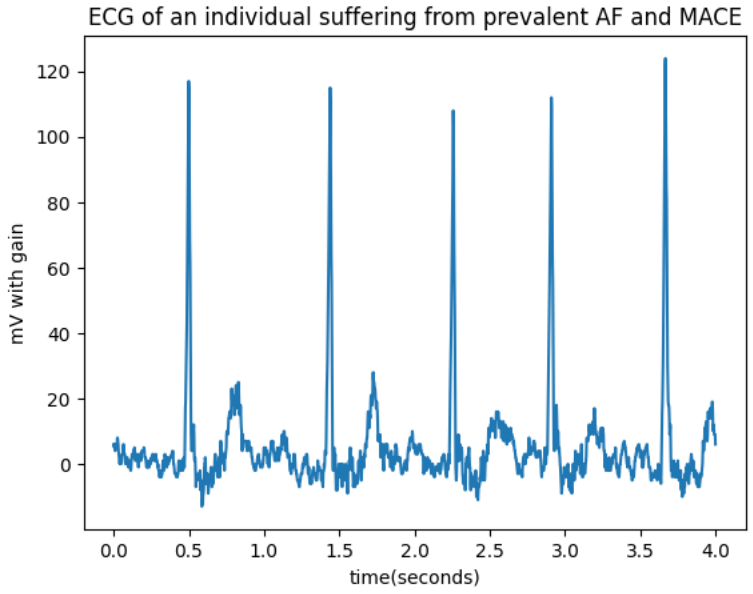

Figure 10: ECG of an individual suffering from prevalent AF and MACE

Data Processing

Small Dataset

The algorithm was organised for a binary classification task (healthy/not healthy) for efficiency, as it could receive both HAF and HMACE datasets. Out of the 66297 raw 1-lead ECGs, 1191 ECGs are follow-up recordings which were discarded.

Consequently, it was important to have the statistics of the population:

Category	**Number of samples**	**Percentage of total population (65106)**
Healthy	57427	88.2%
Prevalent AF	3021	4.64%
Prevalent MACE	3021	4.64%
Incident AF	2778	4.27%
Incident MACE	2663	4.09%

Table 4: Population distribution

Note that some patients suffered or developed both MACE and AF. Prevalent AF and MACE subjects are excluded from the analysis as it is of interest to predict if healthy individuals will suffer from incident AF or MACE. Therefore the 2 datasets are:

	HAF
Class 0: Healthy Individuals	2778
Class I: Individuals with incident AF	2778

Table 5: Class Distribution HAF

	HMACE
Class 0: Healthy Individuals	2663
Class I: Individuals with incident MACE	2663

Table 6: Class Distribution HMACE

The training & validation dataset is balanced by randomly sampling the healthy subjects to extract a number equal to that of the affected patients (Table 6 and Table 7). This way accurate training of the

network is ensured as it has equal access to both classes. The testing dataset is considered as 10% of the new balanced dataset created. The training and validation dataset was further split into 2 parts (detailed in K-Fold Validation Section). The distribution across the training & validation and test datasets are presented in the bar chart below:

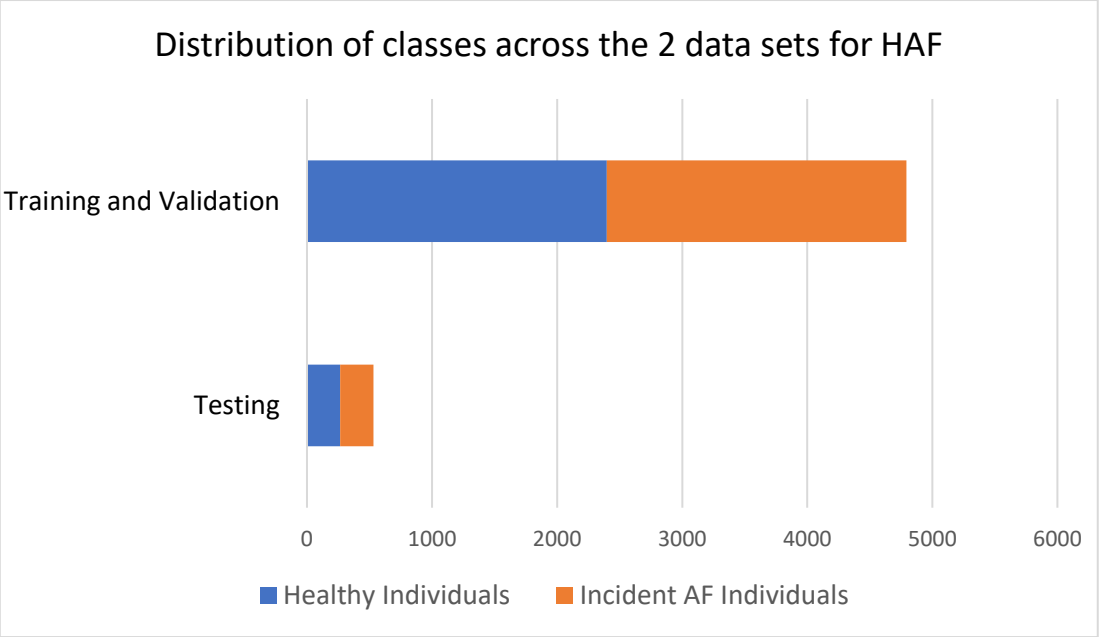

Figure 11: Distribution of AF for the 2 datasets

Figure 12: Distribution of MACE for the 2 datasets

Oversampling (SMOTE) Technique

Deep learning algorithms require large amounts of data for accurate predictions. Due to such a complex time-series input, better results would be expected with an increase in population size. As

regularization techniques were implemented, additional data samples would provide greater precision to the classification algorithm.

Another essential factor is that the network requires access to enough data from both classes. The class imbalance seen in Table 5 is:

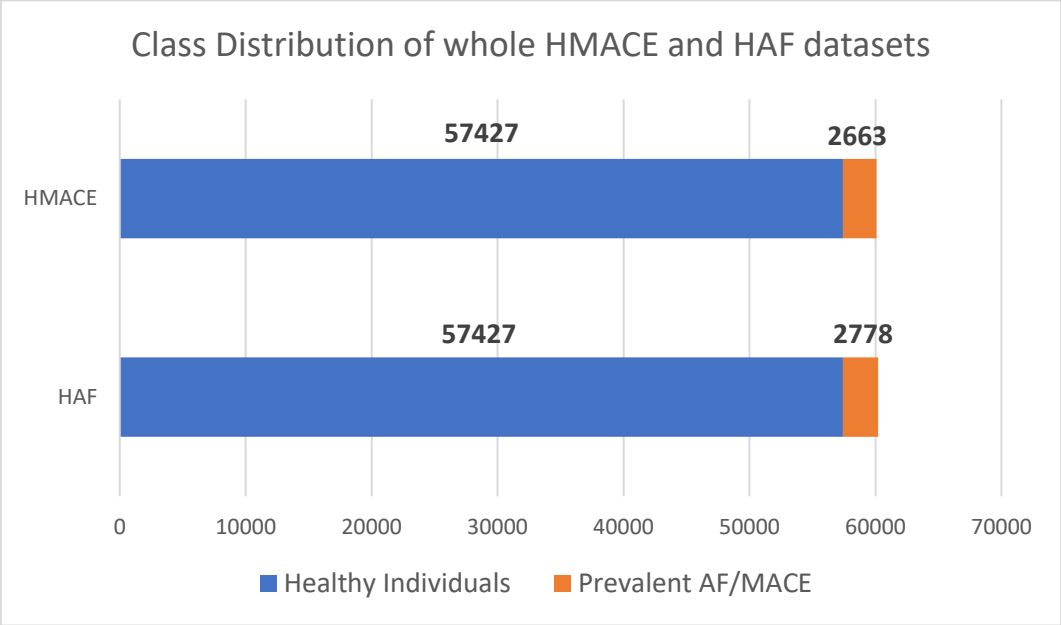

Figure 13: Whole data class distributions of HAF and HMACE

$$\begin{cases} Proportion\ of\ Healthy\ to\ Prevalent\ AF = 1{:}\,0.0463 \\ Proportion\ of\ Healthy\ to\ Prevalent\ MACE = 1{:}\,0.0484 \end{cases}$$

Therefore, by using the whole dataset, the algorithm could predict that an individual will always remain healthy, and be ≈95% accurate. To fix this an oversampling of the minority class (AF/MACE) is required. Oversampling techniques are frequently used to create new data samples of the minority class for the above-mentioned reasons.

One of the most used oversampling techniques is Synthetic Minority Oversampling Technique (SMOTE). SMOTE is efficiently applied in numerical data oversampling with other ECG-AI studies showing pertinent results with this technique.[35] It creates artificial data samples for the minority class to achieve a balanced training dataset. The synthetic data follows an interpolation format:

$$X_{synthetic} = X_i + \lambda(X_j - X_i)$$

Where:

$$\begin{cases} X_{synthetic} = newly\ created\ data\ point \\ X_i = randomly\ selected\ data\ point \\ X_j = randomly\ selected\ k - nearest\ neighbour\ of\ X_i \\ \lambda \in (0,1)\ is\ a\ random\ generated\ variable \end{cases}$$

The k-nearest neighbour algorithm considers the data in an n-dimensional space (n is the number of features of the vector format data) and computes the Euclidean distance to find the k-nearest neighbours.

Considering the general n-dimensional space (ECG with n voltage time steps). Then 2 different ECG samples would be represented as:

$$X_i = (a_1, a_2, a_3, \dots, a_n)$$

$$X_j = (b_1, b_2, b_3, \dots, b_n)$$

The Euclidean distance would be:

$$d(X_i, X_j) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}$$

The SMOTE function computes the k-nearest neighbours for all minority class data points, picks one of the neighbours randomly, generates a new synthetic data point and then restarts the process until the required class distribution is reached.

Large Dataset

For the above-mentioned reasons, a large dataset was created using SMOTE, for a 1:1 proportion in the training dataset. From the data stated in Table 5, the healthy individual cohort was selected in its entirety together with the entire incident AF cohort for the large HAF dataset and with the incident MACE cohort for the large HMACE dataset, respectively. The large class distributions are summarised below:

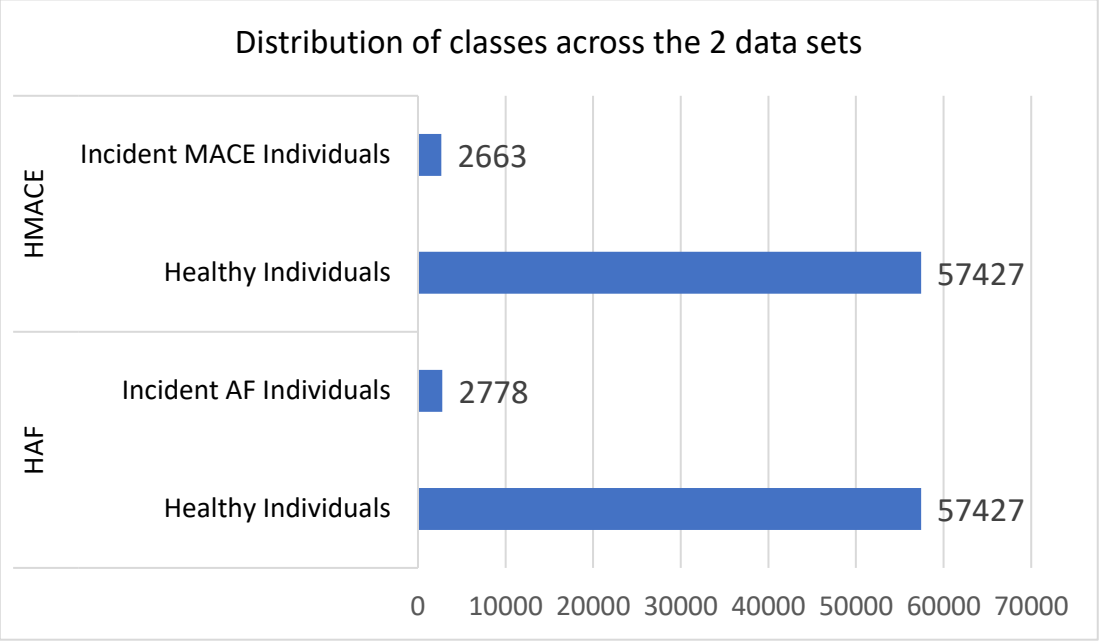

Figure 14: Large Datasets Class Distribution

The above data is split as 20% for testing (keeping the original distribution seen in Figure 13) and 80% training and validation. Oversampling was applied after the training and validation datasets were split as well, only on the training dataset. This will be explained in the K-Fold Cross Validation Section.

Constructed Input Types

This project's main objective is to use exercise stress test data for an accurate classification algorithm. The stages of the recording were displayed in Table 1. Consequently, this project constructs 3 types of inputs described below:

1) 1-Dimensional

This is formed by only selecting the at rest recording. Out of the 15 seconds, the selection was done 2 seconds after the recording onset, for 4 seconds. Therefore, the number of samples is:

$$n = \Delta t * f = 4\ s * 500\ Hz = 2000\ samples$$

The data is in vector format 2000x1. This offers a more compact input, to analyse the efficiency of a smaller sample, compared to other studies which used 10 seconds of recordings.[27] As an arrhythmia, AF, in prevalent form, should be detected with the same precision at rest as in exercise ore recovery, for individuals who suffer from chronic AF. MACE composes of a wider range of affections, which might only show slight abnormalities on Lead I or might only appear in recovery. For example, if MI occurs in the right ventricle (right ventricular infarction) it usually is followed by an ST elevation in Leads II, III and aVF.[36] This format is of interest to assess whether the algorithm could detect even the smallest details in Lead I.

2) 5-Dimensional

This is the most complex format used, also with a 4 seconds length. It has 10000 data points per ECG organised in a 2000x5 matrix:

- Column I: ECG at rest, 2 seconds after the recording onset (1-dimensional input)
- Column II: ECG recorded 30 seconds after the exercise stress test has started (pedalling against a constant workload)
- Column III: ECG recorded 30 seconds after the workload start to increase (linearly increasing workload)
- Column IV: 4 seconds ECG recording at peak exercise
- Column V: 4 seconds ECG recording of the recovery phase

Recordings on different columns were aligned so that the first beat within each epoch occurred at the same time.

Convolutional Neural Network (CNN) Algorithm

Description

CNNs are widely used now in computer vision tasks across the scientific industry, with many applications in biomedical engineering.[37] CNNs are a class of supervised learning algorithms used for classification or prediction problems. They are perfectly suited for grid-like structure analysis (images, time-series data, videos) as they adaptively learn features from the input data. Other potential networks are discussed in the Future Work and Literature Review sections.

The fundamental idea behind a CNN is the use of **convolutional layers** which learn specific features through filters (kernels), reducing the original data size into a more compact form. This makes them particularly useful in image processing, such as an ECG classification problem, by detecting features all over the data sample. CNNs increase efficiency from traditional fully connected machine learning

algorithms as they present fewer parameters to learn.[38] This is attributed to their weight-sharing mechanism of the kernels throughout the input regions.

The convolution operation is an element-wise operation, where the kernel elements are multiplied with the input ones for the respective kernel shape. The output of the operation is called a feature map. An example of a 2D convolution for a 5x5 matrix input with a 3x3 kernel filter is shown below:

$$\begin{pmatrix} 1 & 4 & 8 & 9 & 0 \\ 4 & 2 & 8 & 9 & 2 \\ 2 & 0 & 7 & 5 & 5 \\ 3 & 6 & 0 & 8 & 3 \\ 8 & 3 & 0 & 1 & 9 \end{pmatrix} * \begin{pmatrix} 1 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix} = \begin{pmatrix} -10 & -1 & 16 \\ 4 & -15 & 18 \\ 1 & 0 & -6 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 4 & 8 & 9 & 0 \\ 4 & 2 & 8 & 9 & 2 \\ 2 & 0 & 7 & 5 & 5 \\ 3 & 6 & 0 & 8 & 3 \\ 8 & 3 & 0 & 1 & 9 \end{pmatrix} * \begin{pmatrix} 1 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix} \rightarrow \begin{pmatrix} -10 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

As kernels reduce the input shape more and more throughout the layers, the network has the ability to learn high-level patterns between different spatial parts. The kernels are iteratively updated during the backpropagation algorithm. **Pooling layers** are also highly useful as they further downsample the data, by taking the average (average pooling) or maximum (max pooling) of the values in a local region (specified shape). The convolutional and pooling layers form the **feature extraction part** of the neural network, in which the input is transformed into a feature map to disregard noise and extraneous parts.

Finally, a **linear layer/layers (fully connected)** are used to gather the extracted features and perform mathematical operations via activation functions to produce the output classification. A typical CNN algorithm for image classification is shown below:

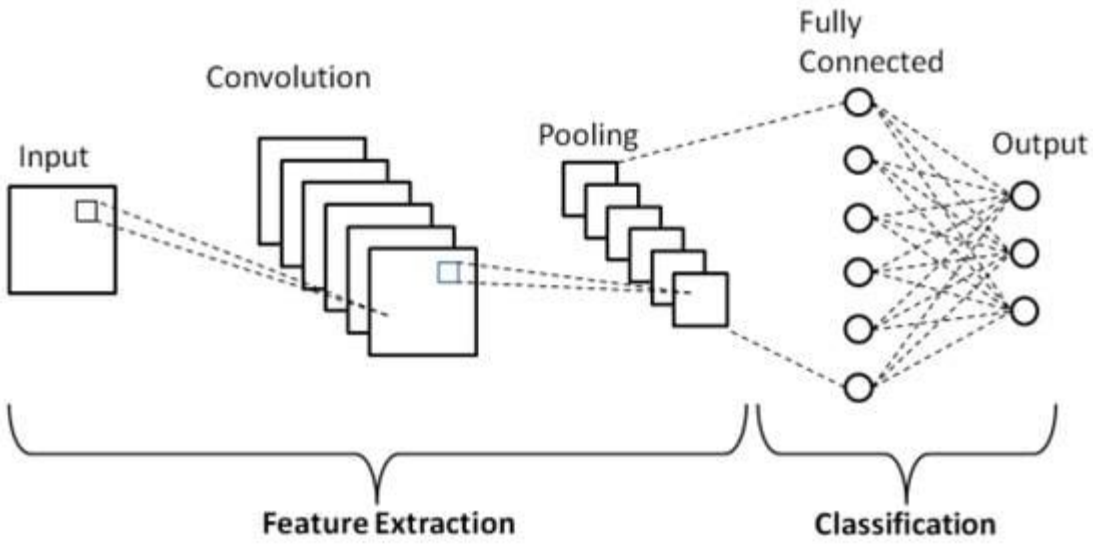

Figure 15: Example CNN architecture[39]

For this project, the ECG data can be interpreted as a column-wise (2000xn) dynamic image. Consequently, a 1D convolution operation is best suited for this type of input. 2D convolution operations have been used in other deep learning 12-lead ECG studies, but this requires a spatial-

-temporal relation in the matrix which only Lead I cannot provide.[40] The full network architecture is detailed in the next section.

Model Architecture and Layers

As a binary classification, it is important to remember the data labels that are used throughout the algorithm for the 2 datasets:

HAF

Class Description	**Labels**
Healthy Individuals	0
Individuals with incident AF	1

Table 7: Class Labels HAF

HMACE

Class Description	**Labels**
Healthy Individuals	0
Individuals with incident MACE	1

Table 8: Class Labels HMACE

1) 1-Dimensional

The network's architecture stemmed from time-series data classification networks and had as a basis in other 1-lead ECG neural networks developed as well as a recent CNN-ECG algorithm which used similar UK Biobank data.[41][42]

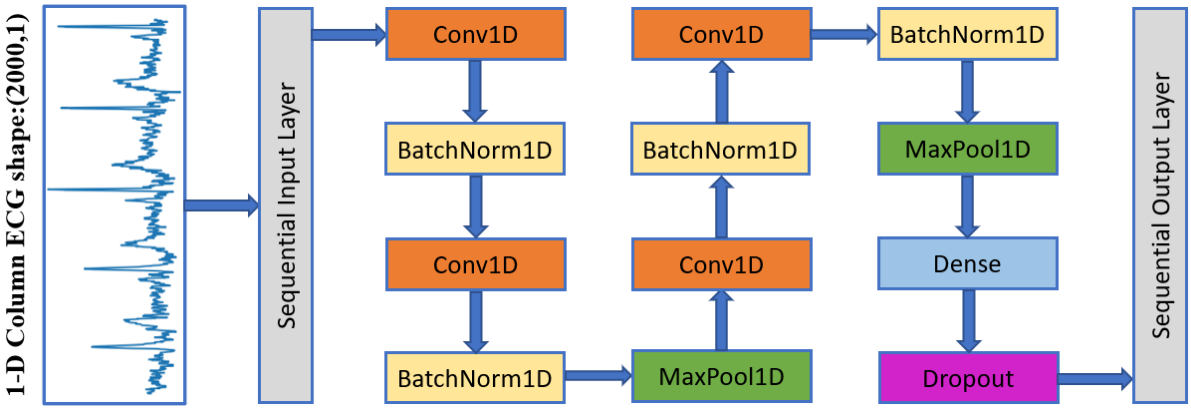

Figure 16: 1-Dimensional Input Network Architecture I

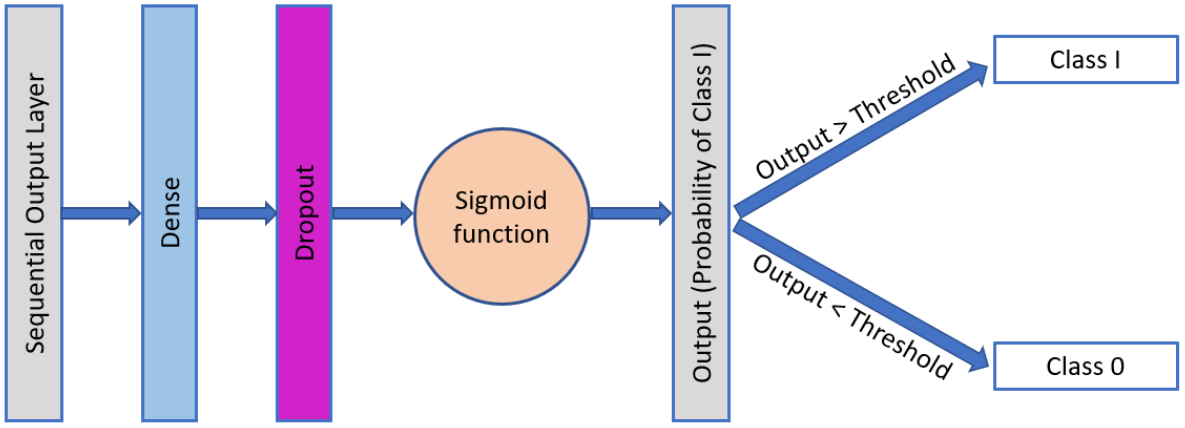

Figure 17:1-Dimensional Input Network Architecture II

The network presents 12 hidden layers in the Sequential Model (Figure 16). The 4 convolutional layers act to reduce the large input size, while increasing the information density with each feature map formed. Each convolutional layer has an increasing number of kernels (32, 64, 128 and 256 respectively). This increases the total number of parameters, creating for a more complex network with better predictability. Kernel sizes of 2 and 3 are kept throughout the network to focus on extracting detailed features from the ECG signal. After each convolutional layer, the output is passed through a Rectified Linear Unit (ReLU) activation function:

$$ReLU(x) = \max(0, x) = \begin{cases} 0 \; for \; x < 0 \\ x \; for \; x > 0 \end{cases}$$

ReLU is a standard activation function in deep learning research. Proposed by Nair and Hinton in 2010, it help with faster computational speeds by introducing sparsity in the convolutional output.[43] Less neurons are consequently activated, reducing the chance of overfitting.

A Batch Normalisation layer is also added after every convolutional layer, so the input is refitted and rescaled. This is done by first calculating the mean and standard deviation of the input batch and normalising it in the following manner:

$$\hat{x}_i = \frac{x_i - \mu_m}{\sqrt{\sigma_m^2}}$$

Where:

$$\begin{cases} x_i = input \; i \; of \; batch \; of \; size \; m \\ \mu_m = mean \; of \; batch \; of \; size \; m \\ \sigma_m = standard \; deviation \; of \; batch \; of \; size \; m \\ \hat{x}_i = internal \; output \end{cases}$$

This internal output is linearised, as all neuron outputs learn the line equation:

$$BatchNorm(\hat{x}_i) = \gamma \hat{x}_i + \beta$$

γ and β are subsequently learned through the iterations in the optimiser. Max Pooling is later applied to further reduce the output dimensions. A fully connected (linear, dense) layer is applied to gather the feature maps in a compact output size out of the sequential model. The last Dropout layer is a regularisation technique that randomly drops neuron outputs to prevent overfitting.

In Figure 17, the 2nd part of the algorithm is presented. Second dense and dropout layers transform the output into a single number. The Sigmoid function, described in the Output section, transform the output into a probability, $p_x \in (0,1)$ from which the classification is produced.

2) 5-Dimensional

The 5-dimensional input, detailed in the Constructed Input types section, is a more complex set of data with 10000 data points. At first, each column was treated separately, with the 1-Dimensional Algorithm applied to all 5, besides the final Sigmoid function. This ensured that features were extracted from each part of the exercise stress test and all 5 columns were considered equally important. All 5 single number outputs were then passed through the architecture below:

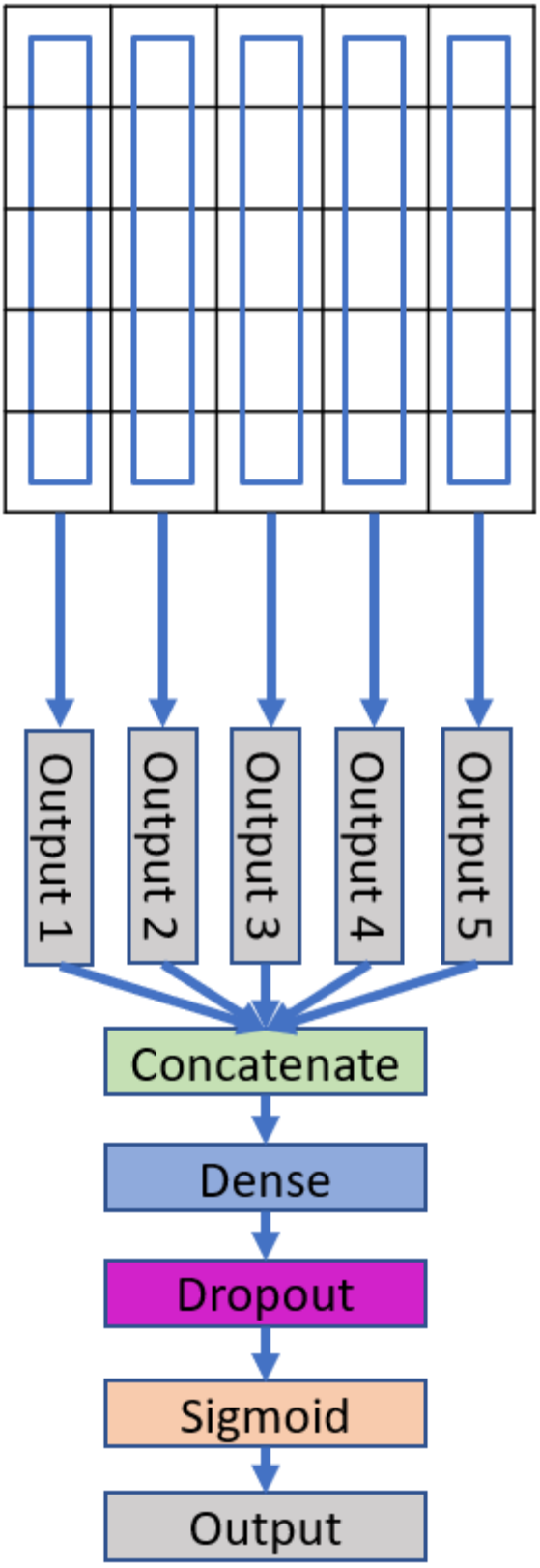

Figure 18: High-level view of 5-dimensional CNN Architecture

Concatenating the outputs, transform them into a single 1x5 vector. A final fully connected layer is used to take into account the products of the columns and produce a final output which is then passed through the Sigmoid function.

Output Type

As described before, the network produces a single output by using the Binary Cross-Entropy Loss function (described in the next section). This output is then passed through a sigmoid function:

$$S(x) = \frac{1}{1 + e^{-x}}$$

This transform it into a number between 0 and 1 which can be interpreted as a probability. The probability represents the likelihood of the respective ECG to be part of Class I (affected class):

$$S(x) = p_x \in (0,1)$$

$$\begin{cases} if\ p_x > threshold \rightarrow Class\ 1 \\ if\ p_x < threshold \rightarrow Class\ 0 \end{cases}$$

Therefore, if the probability is close to 1, the ECG is very likely to be part of Class I. Otherwise, a low probability would infer a closer resemblance to Class 0. These probability values are continuously updated and compared in the loss function to provide better predictions. The threshold was chosen to maximise the Receiver Operating Characteristic (ROC) variables, as explained in the Measuring the Network's Performance section.

Loss Function and Optimiser

Binary Cross-Entropy Loss

The loss function measures the difference between the predicted outputs and the actual labels (0 or 1) to iteratively decrease the error. The loss function was chosen as the Binary Cross-Entropy Loss, also referred to as the Logarithmic Loss[44]:

$$BCE_{Loss} = -\frac{1}{N}\sum_{i=1}^{N} -(y_i * \log p_i + (1 - y_i) * \log(1 - p_i))$$

Where:

$$\begin{cases} y_i = label\ of\ sample\ i\ (either\ 0\ or\ 1) \\ p_i = predicted\ probability\ of\ sample\ i \\ N = batch\ size \end{cases}$$

This is designed to work for a singular output representing the class 1 probability. If $p_i = 1$, that is perfectly fitted to class 1, then the $BCE_{Loss} = 0$. As the probability nears zero for a label of 1, the logarithm will reach low negative values, which will result in a high loss.

Stochastic Gradient Descent (SGD) Optimiser

An optimisation algorithm is meant to update the weight and biases of a neural network through the loss function. This increases the performance of the network by decreasing the loss function through the epochs to reach a global minimum. Therefore, the predicted labels get closer to the actual labels. SGD is derived from the classical gradient descent algorithm with the main difference is that it splits the training dataset into "mini-batches" to be faster and memory-efficient. It has become an industry

standard, together with the newer Adam optimiser for image processing techniques.[45] Comparisons of both SGD and Adam were done and the different network architectures showed better stability and generalisation with SGD.

Regularisation techniques

For best performance, regularisation techniques such as L1 (Lasso Regularisation) and L2 (Ridge Regularisation) are implement to avoid overfitting. A penalty term is added to the loss function to prevent it being over dependent on certain features which will show to have higher weights. Below the L2 regularisation method is presented:

$$Loss\ Function_{L2} = -\frac{1}{N}\sum_{i=1}^{N} -(y_i * \log p_i + (1 - y_i) * \log(1 - p_i)\ ) + \lambda \sum_{i=1}^{N} W^2$$

Where:

$$\begin{cases} \sum_{i=1}^{N} W^2 = sum\ of\ all\ model\ parameters\ (neuron\ weights) \\ \lambda \in (0,1)\ weight\ decay\ parameter \end{cases}$$

$$\sum_{i=1}^{N} W^2 = sum\ of\ all\ model\ parameters\ (neuron\ weights)$$

This was included in the script in the Adam optimiser initialisation through a weight decay parameter $(\lambda = 0.05)$.

K-fold Cross Validation

A powerful tool that is used in advanced neural networks to prevent overfitting is K-fold Cross Validation.[46] After a network is trained it should have access to an external dataset in order to test its validity and further update the model. This involves using a validation dataset before the actual test set (detailed in the Small and Large Dataset sections).

K-fold cross validation is a more advanced method that, at first, splits the dataset into a training and testing only dataset. The training dataset is separated further in a loop in k folds. The first k-1 folds are used for training and the last fold is used to validate the model. This allows the model to be tested constantly on different parts of the dataset. The procedure is exemplified in Figure 16. This is all done while keeping the same distribution (Stratified K-fold Validation) of classes in all splits to maintain the algorithm's unbiased view of the samples. SMOTE is applied after a fold is created and only on the k-1 folds in the current training dataset. As a result, validation and testing only contain real data, so the network's accuracy is impartial to the 2 classes.

A 5-fold Stratified Cross Validation was employed for this project. For a bigger k, the number of minority class examples (AF or MACE) would be too few in each fold. The oversampling would lead to almost identical class I samples, which will bias the algorithm in its prediction.

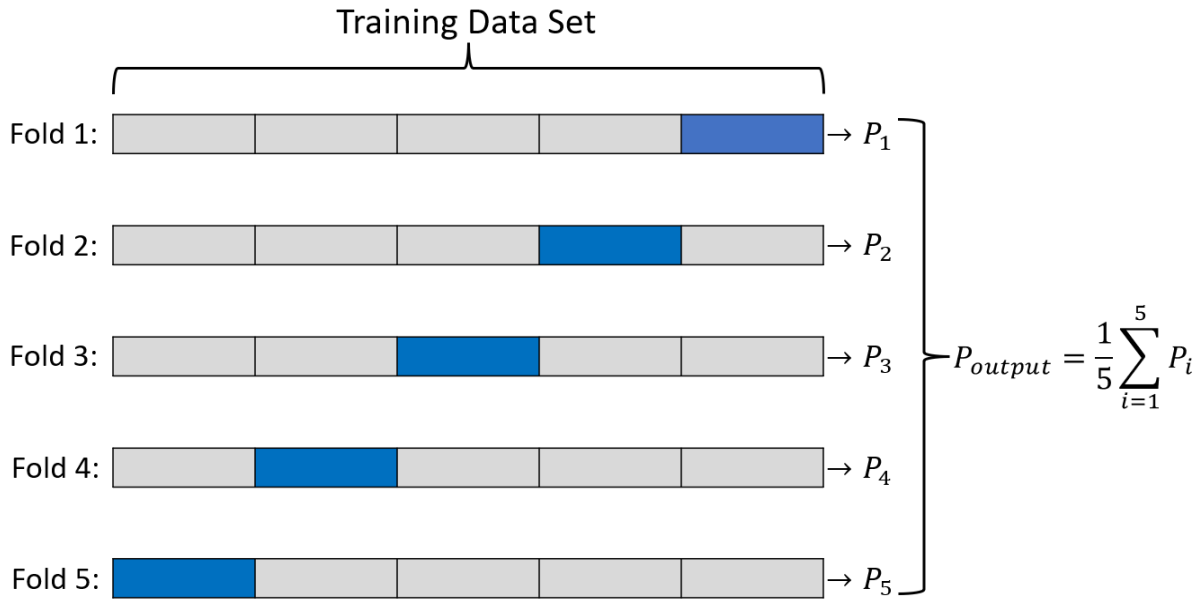

Figure 19: 5-fold Cross-Validation Description

Use of Linux and UCL Research Computing Services

A part of the analysis pertaining to the statistical data was done using UCL's Research Computing Services on Myriad. This allowed for a higher RAM capacity and faster speeds. The code was uploaded on Myriad and run by a Linux Bash script include in the Appendix.

Measuring the Network's Performance

For the purpose of quantifying the results of the CNN, it's Receiver Operating Characteristic (ROC) curve was constructed. The ROC is seen as the gold standard in reporting the results of a binary classifier.[47] In medical diagnostic tools, it is important to not rely on the total accuracy of the predicted results, but also to the True Positive Rate (TPR) and False Positive Rate (FPR):

$$TPR = \frac{true\ positive\ cases\ identified}{total\ positive\ cases}$$

$$FPR = \frac{false\ positive\ cases\ identified}{total\ negative\ cases}$$

These values can also be defined with the terms Sensitivity and Specificity in the following manner:

$$\begin{cases} Sensitivity = TPR \\ Specificity = 1 - FPR \end{cases}$$

The ROC is reported as a plot representing the relationship between TPR and FPR at different thresholds. The Area under the ROC curve, AUROC, is considered the standard numerical value to assess the dichotomous classifier. An Area of 1 would mean a perfect classification. An Area of 0.5 naturally leads to a completely random classification. The random line (AUROC = 0.5) is represented in Figure 20 with green.

In the Output type section, it was described that the result is a probability of the ECG sample being part of Class I. If the probability is larger than a threshold, then it is attributed to Class I, otherwise it is Class 0. Different thresholds yield different a different TPR and FPR, so it is important to maximise

the true positive cases while minimising the false positive cases. For this, Youden's Index was maximised:

$$J = Sensitivity + Specificity - 1$$

$$J = TPR - FPR$$

An ROC curve is reported below with the point where Youden's Index was maximised for reference:

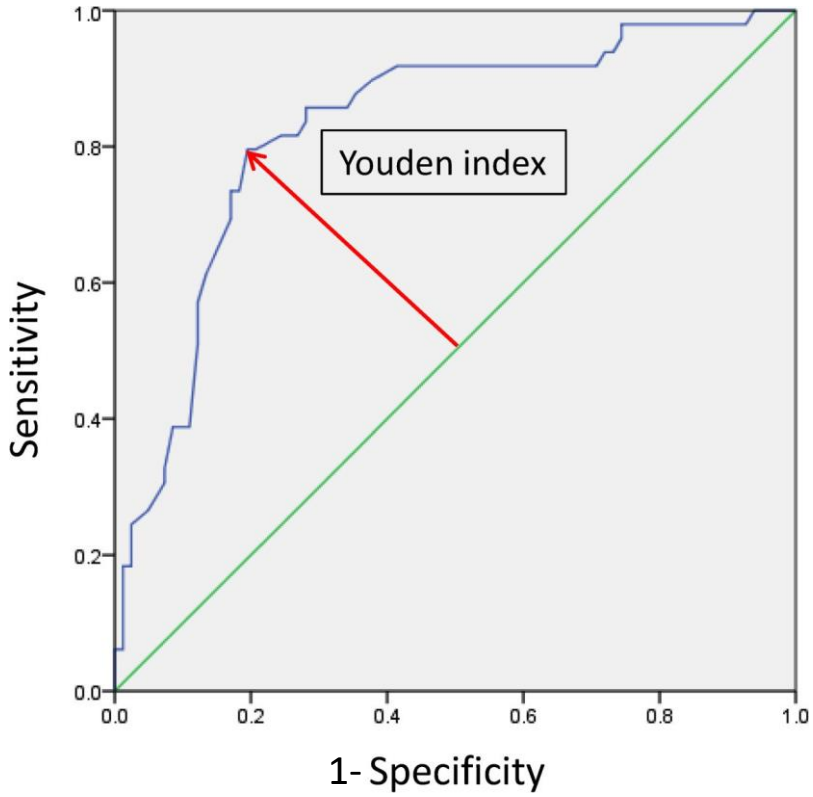

Figure 20: Informative ROC Curve and Youden's Index[48]

Consequently, the threshold for maximising J was computed with the code attached in the Appendix.

‘

Results and Discussion

Small Dataset

For both the 1-Dimensional and 5-Dimensional inputs in the small dataset, the following hyperparameters were chosen through empirical evaluation:

Hyperparameter	**Value**
No. of Epochs	20
Batch Size	32
Learning Rate	0.001
Weight Decay	0.05
No. of Folds	5

Table 9: Hyperparameters of the network for the Small Dataset

1-Dimensional

The following ROC curves were obtained from the 1-dimensional inputs for the HAF and HMACE datasets:

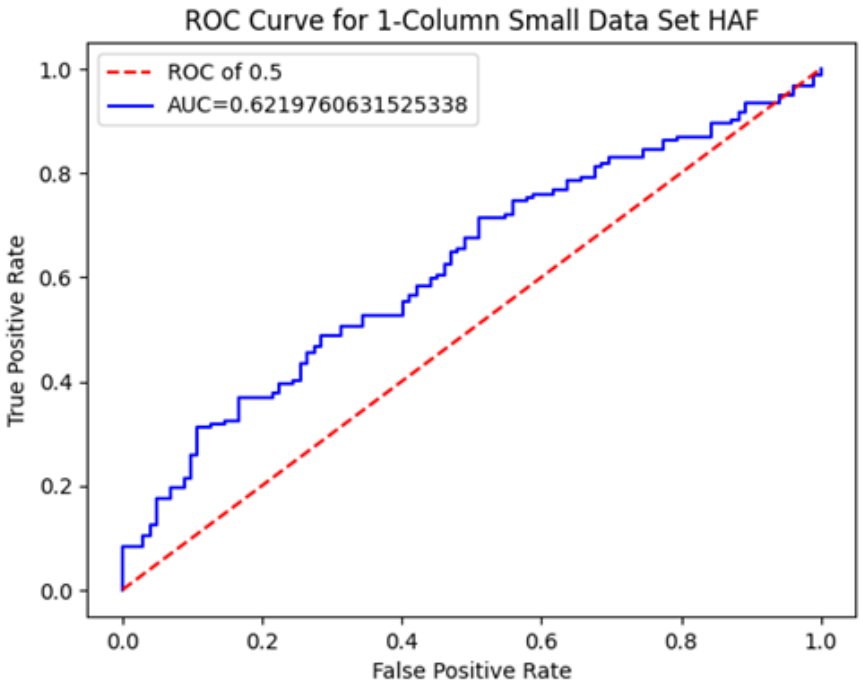

Figure 21: 1-Dimensional Small Dataset (HAF) ROC Curve

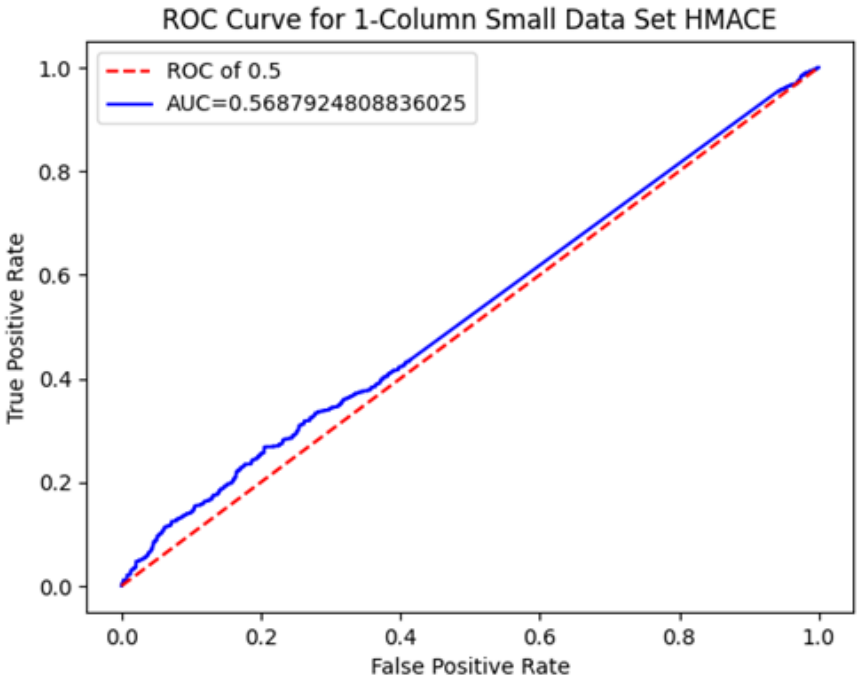

Figure 22: 1-Dimensional Small Dataset (HMACE) ROC Curve

The evaluation metrics are also reported in the table below:

	HAF	**HMACE**
Dataset Size	5556	5326
AUROC	0.622	0.569
Threshold	0.55	0.58
TPR (cut-off point)	0.714	0.345
FPR (cut-off point)	0.510	0.286

Table 10: HAF and HMACE Small Dataset Results for 1-Dimensional Input

For this input and datasets, the network was trained for 20 epochs with a standard batch size of 32 on each fold to prevent overfitting. As expected, the ROC curve of the HAF is considerably better than that of the HMACE dataset. The TPR indicates that 71% of the positive cases of AF were identified, although the network seemed to also output many false positives. Therefore, for the small HAF dataset, the network has good sensitivity, but average specificity. These results can be explained by the more facile diagnostic of prevalent AF. As detailed in the Literature Review section, AF is usually represented on the ECG signal by the absence of the P-wave (atrial depolarization) and an irregular R-R interval. MACE is a composite endpoint of a multitude of different cardiovascular affection that could be exhibited throughout the PQRST wave. Consequently, the network didn't receive sufficient information for the MACE dataset to capture all possible cardiovascular disfunctions that could appear. In addition, it is more relevant to assess MACE considering the exercise stress test and the recorded recovery period.

The 1-Dimensional input was comprised of at-rest recordings in which chronic AF is always present in the affected individual.

The 230 extra samples (Table 10) of the HAF dataset could've also provided a minor advantage in the AUROC results as the network had approximately 4.3% more information available.

Overall, lower AUROC results were expected for the small dataset. Nevertheless, the AUROC of 0.622 for the HAF dataset shows the prediction potential for incident AF, years before medical check-ups are able to pick up certain signals.

5-Dimensional

The following ROC curves were obtained from the 5-dimensional inputs for the HAF and HMACE datasets:

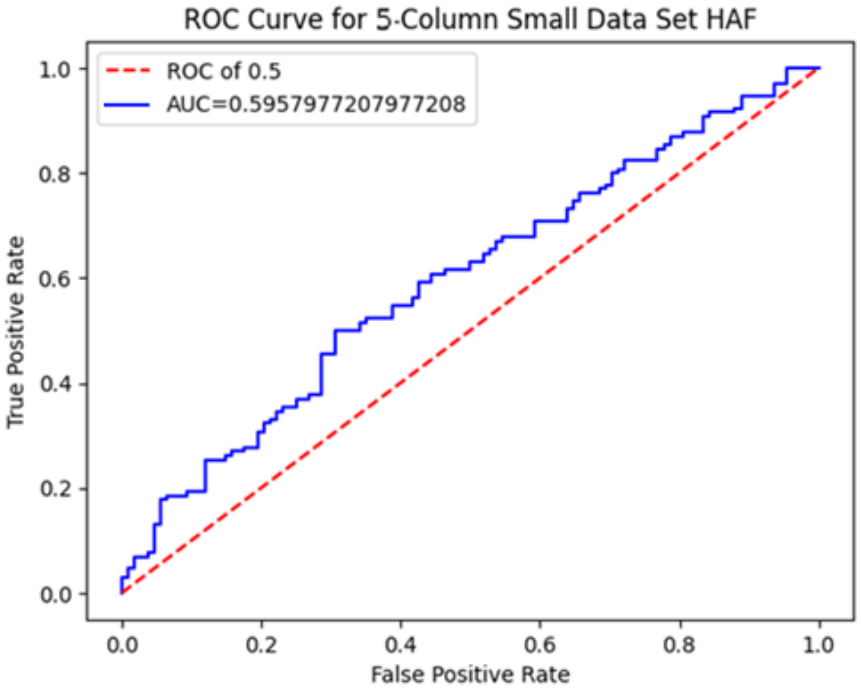

Figure 23: 5-Dimensional Small Dataset (HAF) ROC Curve

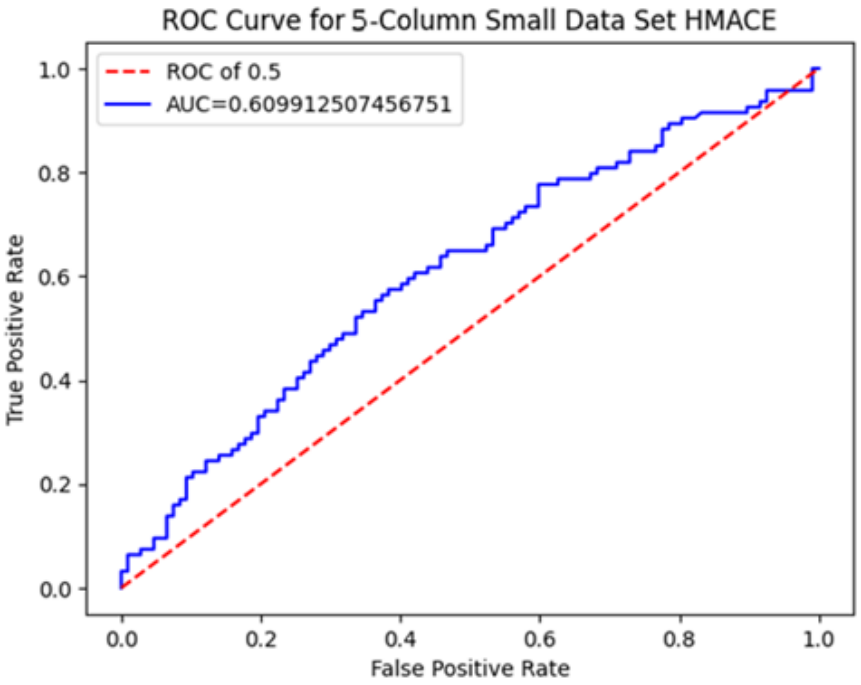

Figure 24: 5-Dimensional Small Dataset (HMACE) ROC Curve

The metrics are also reported in the table below:

	HAF	**HMACE**
Dataset Size	5556	5326
AUROC	0.596	0.610
Threshold	0.59	0.58
TPR (cut-off point)	0.644	0.544
FPR (cut-off point)	0.454	0.343

Table 11: HAF and HMACE Small Dataset Results for 5-Dimensional Input

The idea to incorporate multiple parts of the exercise stress test into a prediction algorithm was to test if improvements in the accuracy algorithm are observed, especially for the MACE dataset. It can be seen that, although the AF prediction remained relatively constant, with a similar FPR-TPR relation, the MACE results improved considerably (0.569 to 0.610). The recovery and exercises phases assisted in the prediction procedure and this provides additional motivation to understand the performance of the algorithm on the large dataset.

Large Dataset

For both the 1-Dimensional and 5-Dimensional inputs in the large dataset, the following hyperparameters were chosen through empirical evaluation:

Hyperparameter	**Value**
No. of Epochs	10
Batch Size	64
Learning Rate	0.001
Weight Decay	0.05
No. of Folds	5

Table 12: Hyperparameters of the network for the Small Dataset

1-Dimensional

The following ROC curves were obtained from the 1-dimensional inputs for the HAF and HMACE datasets:

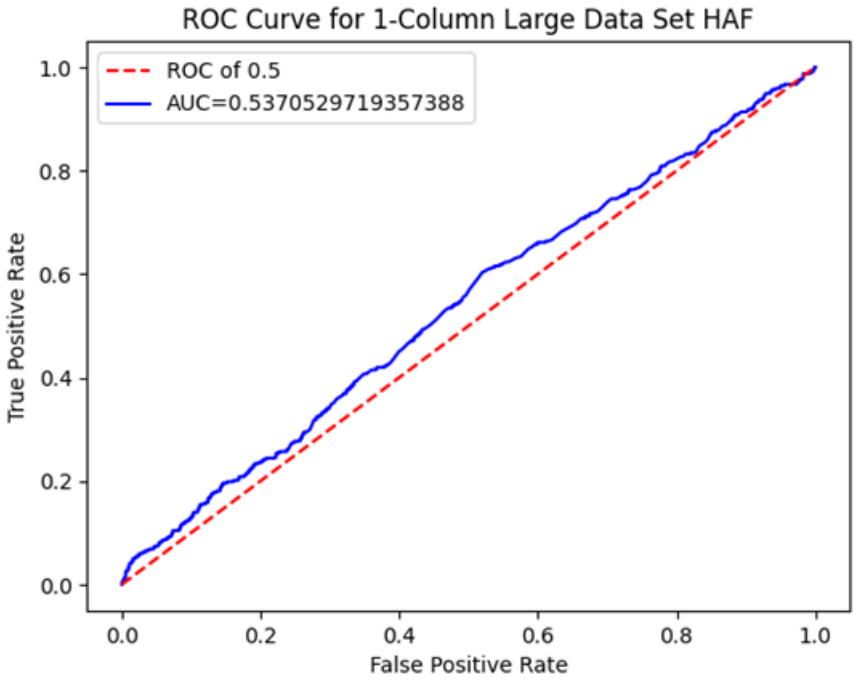

Figure 25: 1-Dimensional Large Dataset (HAF) ROC Curve

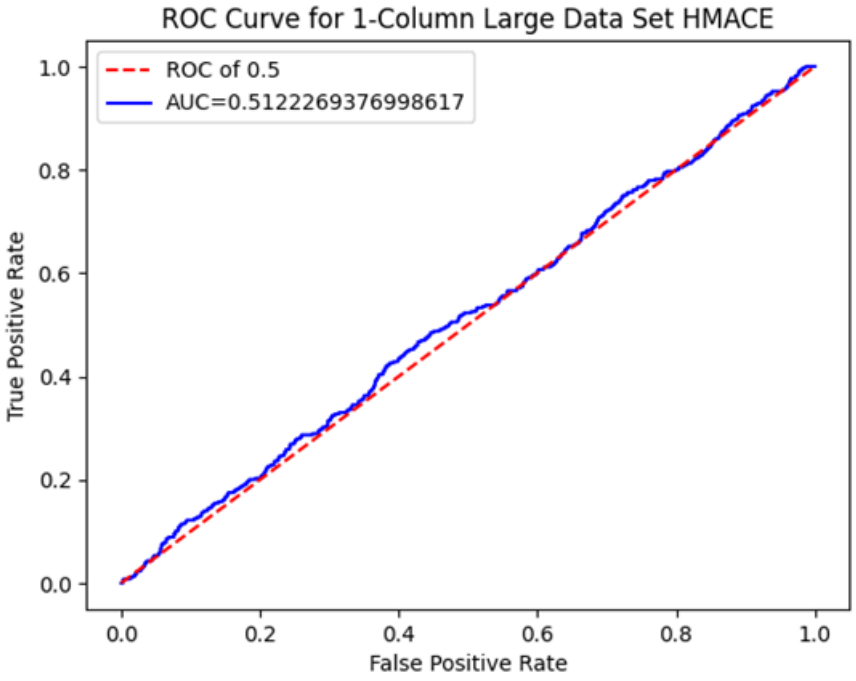

Figure 26: 1-Dimensional Large Dataset (HMACE) ROC Curve

The metrics for the 1-Dimensional Large Dataset are reported in the table below:

	HAF		**HMACE**	
Original Dataset Size	60090		60205	
Original Dataset Distribution	57427 (H)	2663 (AF)	57427 (H)	2778 (MACE)
AUROC	0.537		0.512	
Threshold	0.55		0.55	
TPR (cut-off point)	0.603		0.467	
FPR (cut-off point)	0.520		0.437	

Table 13: HAF and HMACE Large Dataset Results for 1-Dimensional Input

5-Dimensional

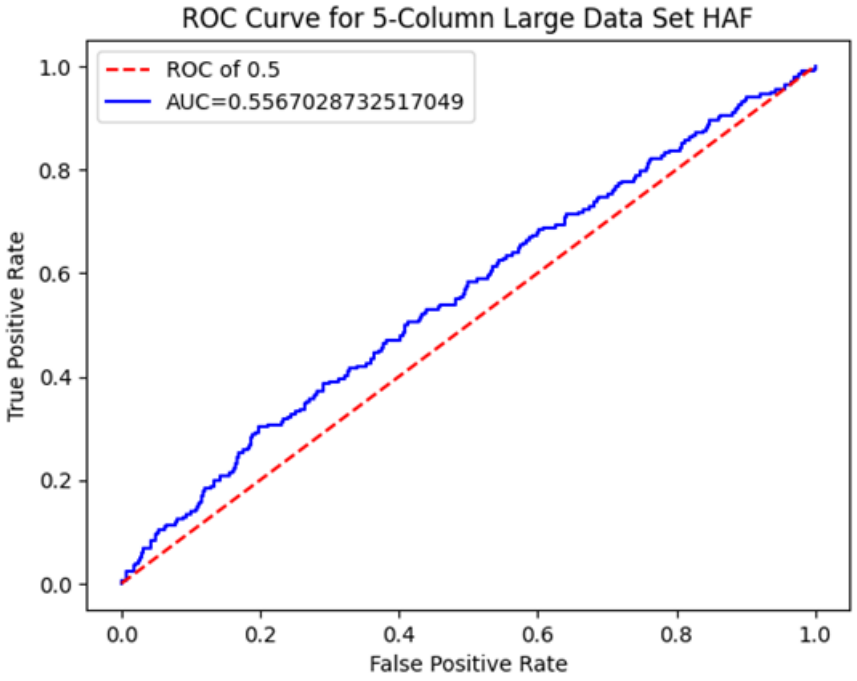

Figure 27: 5-Dimensional Large Dataset (HAF) ROC Curve

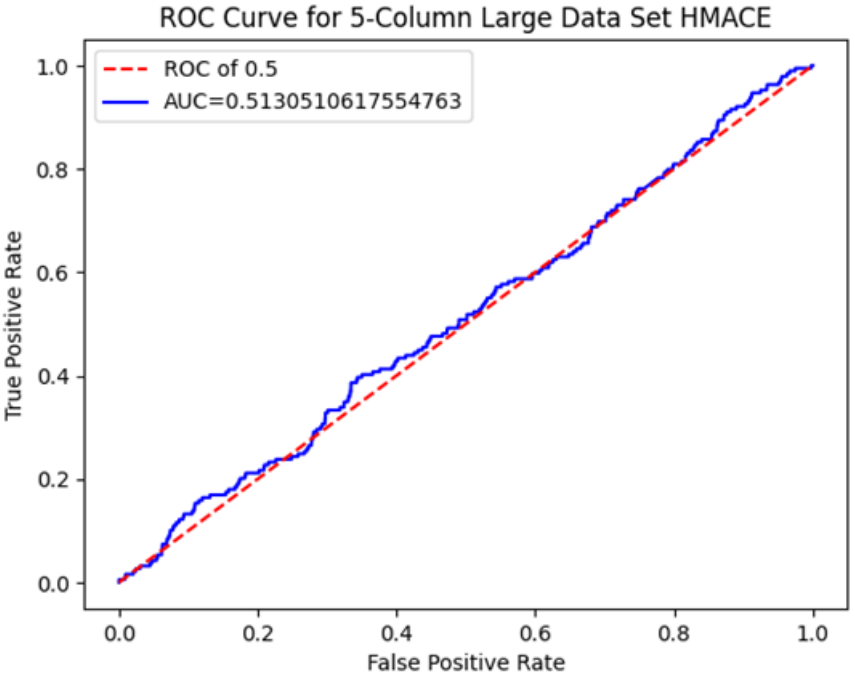

Figure 28: 5-Dimensional Large Dataset (HMACE) ROC Curve

The metrics for the 5-Column Large Dataset are reported in the table below:

	HAF		HMACE	
Original Dataset Size	60090		60205	
Original Dataset Distribution	57427 (H)	2663 (AF)	57427 (H)	2778 (MACE)
AUROC	0.557		0.513	
Threshold	0.55		0.55	
TPR (cut-off point)	0.472		0.439	
FPR (cut-off point)	0.411		0.408	

Table 14: HAF and HMACE Large Dataset Results for 5-Dimensional Input

The analysis for the large dataset was conducted to increase the accuracy of the ROC curve results. In Table 12 the Original Dataset distribution is reported, with 57427 healthy individuals, 2663 AF affected and 2778 suffering from incident MACE, respectively. The SMOTE oversampling (Oversampling (SMOTE) and Large Dataset Sections) was applied inside the training dataset, only on the training part of the fold.

The results are poor overall for both 1-Dimensional and 5-Dimensional inputs, with the problem being tracked to the SMOTE technique application. The SMOTE calculated the K-nearest neighbours for the ECG samples without taking into account for the different patterns occurring from one individual to another. It is clear that the 4 seconds recorded will capture different heartbeat variations in both time occurrence and number. Consequently, when using the Euclidean distance to calculate the neighbours of the sample, although a close match was found, the recorded datapoints were at different points of the PQRST complex. A figure below represents this more clearly:

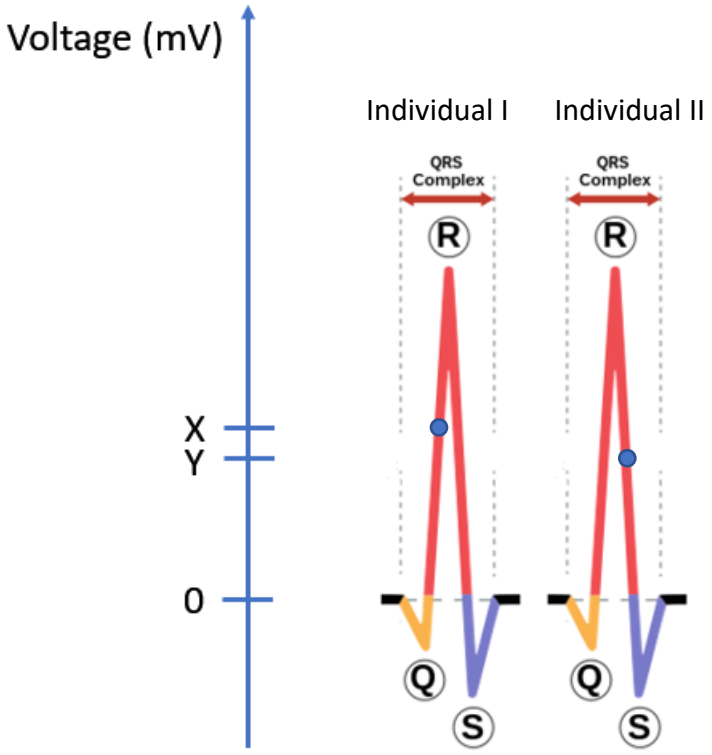

Figure 29: QRS complex of two different samples [16]

The ECG wave can be defined as a function of voltage with time. Due to the start of the recording encompassing different parts of the complex, an erroneous interpolation may occur. Although the individuals above have a similar measured voltage $(X \cong Y)$, the timings are totally different: one occurring at the start of the ventricular depolarisation and one at the end of it.

This introduced additional noise to the already unfiltered ECG samples. Consequently, the network learns patterns that do not exist in real samples. This was confirmed by widely different results in the training loss (where oversampling was conducted) and the validation loss. The validity of other oversampling techniques is discussed in the next section.

Overall Observations and Comparison with AI Industry Results

With regards to the knowledge involved in this project, there are only a few studies which have developed predictive models for AF and none for MACE. Khurshid et al. developed a CNN model for 5-year incident AF prediction.[49] The output raw probabilities of the deep learning model became the inputs to 3 Cox Regression Models. The model was also tested on the UK Biobank for validation purposes and it achieved a larger 0.746 AUROC. It is to be noted that the datasets used by Khurshid et al. are formed of 12-lead ECGs, not 1-lead as is the case here, and they accounted for the time of the recording. The time-series included in the study was also 150% larger, with 5000 voltage measurements compared to only 2000. Most studies for ECG-based prevalent detection also used a 10s input (500Hz frequency) to capture more information, but the validity of the input size was never discussed.

Other studies have assessed short-term predictions for AF. Raghunath et al. published an article in 2021 which describes the design of a deep CNN which achieved an AUROC of 0.85 on 1-year incident AF prediction.[50] The samples were formed of 10s 12-lead recordings with a very numerous cohort, 1.6 million ECGs from 430000 individuals.

Regarding the work completed in this project, some noteworthy improvements are worth mentioning. The AUROC of 0.622 for 1-Column AF prediction on 1-lead ECGs showcases the potential of applying convolutional algorithms to medical prediction problems and further improvements are strongly advised.

SMOTE was not adequate for creating statistical artificial data in this case. Too much noise was added to the new samples, increasing the difficulty of prediction. A viable technique would be Random Oversampling. This would involve copying minority class ECGs in training. The main issue would be that for a higher oversampling rate (>80%) this would encourage the network to memorise the training data, rather than learning its features. This would cause overfitting problems, as the algorithm won't be able to make accurate predictions on the test dataset.

A 2020 article published by Zhu et al. detailed the limitations of oversampling techniques related to imbalanced time series datasets. A method that differs from oversample-based interpolation (SMOTE) is proposed called OHIT (Oversampling High-Dimensional Imbalanced Time Series).[51] OHIT generates structure-preserving synthetic samples by measuring the samples covariance in the class, instead of comparing local characteristics of specific samples.

Another solution for dealing with imbalanced datasets instead of oversampling would be to use a loss function that penalises wrong minority class predictions. This can be achieved by assigning weights to the currently used BCE loss so that it penalises misclassification of the minority class by a larger factor

than for a majority misclassification. Other pre-implemented loss functions, such as the Focal Loss, have been successfully implemented in class imbalance problems.[52]

Conclusion

Future Work

Other types of architectures should be tried on a "dynamic image" such as the ECG. Architectures such as LSTM could potentially better extract time-series data features, especially in a hybrid CNN-LSTM network. A feedback connection (Bi-LSTM) for the algorithm to perceive if it extracted the relevant features in its training should also be utilised. The risk quantification could also be more precise. Instead of doing a binary classification, a multiclass with each class representing different risk levels and timelines of development would be more accurate. This would also split MACE predictions into several, more facile, diagnostics.

A new class of deep learning models that should be implemented are called transformers. Transformers are now widely researched in the field of computer vision for their state-of-the-art performance on sequential input types of tasks. Proposed in a 2017 paper by Vaswani et al. they use an encoder-decoder structure with attention-based layers.[53] The attention-based blocks allow the model to split the input and weight it to learn to predict only the important parts, surpassing complex LSTM architectures for classification tasks.

On a lower level, the algorithm would benefit from a GridSearch implementation. GridSearch checks different values of the network's hyperparameters (learning rate, batch size etc.) for the best output performance. Denoising the data by applying high-pass and low-pass filters would enhance the algorithm's ability to extract the relevant features. Finally, additional data will always be valuable in improving the performance metrics of the network. Consequently, it would be recommended to use other datasets from aforementioned studies and combine them into a unified dataset.

Objectives Completed

This project has achieved the following objectives:

- Successfully pre-processed 66297 1-lead ECGs to form the training and validation datasets
- Created synthetic data via the SMOTE technique
- Developed 2 deep CNNs for the 1-Dimensional and 5-Dimensional inputs
- Tested the validity of the model by plotting the ROC and analysing the network's specificity and sensitivity on the different datasets

The extraneous data was deleted from the 66297 ECGs and the rest was split into training-validation and testing datasets. Two self-designed architectures were used which made use of the Cross-Validation technique, splitting the data into 5 folds and applying SMOTE inside the folds. The proven AUROC of the HAF on the 1-Column (Small Dataset) and the increase in AUROC from the 1-column to the 5-column analysis (Small Dataset) for HMACE prove the potential of ECG-enabled cardiovascular disease classification algorithms. Further improvement in terms of the architecture and oversampling technique employed is needed to achieve better results.

Bibliography

1. M. John Lever, The cardiovascular system, *Biomaterials, Artificial Organs and Tissue Engineering*, volume 9, Science Direct, Woodhead Publishing Series in Biomaterials, pages 90-96, doi.org/10.1533/9781845690861.2.90, 2005
   Available at: https://www.sciencedirect.com/science/article/pii/B9781855737372500092
2. Roland X. Stroobandt, S. Serge Barold, Alfons F. Sinnaeve, *ECG from Basics to essentials step by step*, 2015, ISBN 9781119066422, Publisher: Wiley Blackwell
3. Richard E. Klabunde, 2017, January 31, Cardiac electrophysiology: normal and ischemic ionic currents and the ECG, *The Adv Physiol Educ* 41: 29–37, doi.org/10.1152/advan.00105.2016, 2017
   Available at: https://journals.physiology.org/doi/full/10.1152/advan.00105.2016
4. NHS, *Atrial Fibrillation NHS choices*. NHS. Available at: https://www.nhs.uk/conditions/atrial-fibrillation/.
5. Cables and Sensors, *12 lead ECG Placement Guide: Cables & Sensors*, *Cables and Sensors*. Available at: https://www.cablesandsensors.com/pages/12-lead-ecg-placement-guide-with-illustrations.
6. Jonathan Mant, David A Fitzmaurice, F D Richard Hobbs, Sue Jowett, Ellen T Murray, Roger Holder, Michael Davies, Gregory Y H Lip, *Accuracy of diagnosing atrial fibrillation on electrocardiogram by primary care practitioners and interpretative diagnostic software: analysis of data from screening for atrial fibrillation in the elderly (SAFE) trial*, BMJ, 2007, August, doi: 10.1136/bmj.39227.551713.AE,
   Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1952490/#
7. Zachi I. Attia, Suraj Kapa, Francisco Lopez-Jimenez, Paul M. McKie, Dorothy J. Ladewig, Gaurav Satam, Patricia A. Pellikka, Maurice Enriquez-Sarano, Peter A. Noseworthy, Thomas M. Munger, Samuel J. Asirvatham, Christopher G. Scott, Rickey E. Carter & Paul A. Friedman, Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram, *Nature Medicine* volume 25, pages 70-74, ISSN 1546 -170X (online), ISSN 1078-8956 (print), 2019, January, doi: 10.1038/s41591-018-0240-2,
   Available at: https://doi.org/10.1038/s41591-018-0240-2
8. Julien Oster, Jemma C Hopewell, Klemen Ziberna, Rohan Wijesurendra, Christian F Camm, Barbara Casadei and Lionel Tarassenko, Identification of patients with atrial fibrillation: a big data exploratory analysis of the UK Biobank, *Physiological Measurement*, Volume 41, Number 2, 2020, March, doi: 10.1088/1361-6579/ab6f9a,
   Available at: https://doi.org/10.1088/1361-6579/ab6f9a
9. Byoung Geol Choi, Seung-Woon Rha, Seong Gyu Yoon, Cheol Ung Choi, Min Woo Lee and Suhng Wook Kim, Association of Major Adverse Cardiac Events up to 5 Years in Patients With Chest Pain Without Significant Coronary Artery Disease in the Korean Population, *Journal of the American Heart Association*. Vol 8, No 12, 2019, June,
   doi: doi.org/10.1161/JAHA.118.010541,
   Available at: https://doi.org/10.1161/JAHA.118.010541
10. Constance Wou, James Crompton, Mark Ashworth, Helen Williams, Hiten Dodhia, Managing stroke risk in patients with atrial fibrillation: a cross-sectional analysis of socio-demographic inequalities in a London borough, *Journal of Public Health*, Volume 44, Issue 2, June 2022, Pages e241–e248, https://doi.org/10.1093/pubmed/fdac004

11. *The CVD Challenge in England*, *British Heart Foundation*. Available at: https://www.bhf.org.uk/for-professionals/healthcare-professionals/data-and-statistics/the-cvd-challenge/the-cvd-challenge-in-england.
12. Alkhatib, A. (2019) *How to get the most of your medical consultation*, *How to Get the Most of Your Medical Consultation*. Available at: https://thedoctorweighsin.com/how-to-get-the-most-of-your-medical-consultation/.
13. *Electrocardiogram (ECG or EKG)* (2022) *www.heart.org*. Available at: https://www.heart.org/en/health-topics/heart-attack/diagnosing-a-heart-attack/electrocardiogram-ecg-or-ekg.
14. *How to do a SWOT analysis in Healthcare: Clearpoint strategy* (no date) *ClearPoint Strategy*. Available at: https://www.clearpointstrategy.com/blog/swot-analysis-in-healthcare.
15. Borad, A. (2019) *How to develop machine learning applications for business*, *Product Engineering Services*. Available at: https://www.einfochips.com/blog/how-to-develop-machine-learning-applications-for-business/.
16. A. Atkielski, "Schematic diagram of normal sinus rhythm for a human heart as seen on ecg," January 2007.
Available at: https://commons.wikimedia.org/wiki/File:SinusRhythmLabels.svg
17. Jamieson M. Bourque, M.D., M.H.S.[*†] and George A. Beller, Value of Exercise Stress Electrocardiography for Risk Stratification in Patients With Suspected or Known Coronary Artery Disease in the Era of Advanced Imaging Technologies, *JACC Cardiovasc Imaging.* 2015 Nov; 8(11): 1309–1321.,
doi: 10.1016/j.jcmg.2015.09.006
Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4646721/
18. Yoriko Heianza, RD, PhD,[1] Wenjie Ma, MD, MS,[2] JoAnn E. Manson, MD, DrPH,[2,4] Kathryn M. Rexrode, MD, MPH,[4] and Lu Qi, MD, PhD, Gut Microbiota Metabolites and Risk of Major Adverse Cardiovascular Disease Events and Death: A Systematic Review and Meta-Analysis of Prospective Studies*, J Am Heart Assoc.* 2017, July; 6(7): e004947. PMID: 28663251,
doi: 10.1161/JAHA.116.004947
Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5586261/
19. Elliott Bosco, Leon Hsueh, Kevin W. McConeghy, Stefan Gravenstein and Elie Saade, Major adverse cardiovascular event definitions used in observational analysis of administrative databases: a systematic review, *BMC Med Res Methodol*., 2012, November, doi: 10.1186/s12874-021-01440-5,
Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8571870/
20. Alexander Muacevic and John R Adler, Ishan Poudel,[1] Chavi Tejpal,[2] Hamza Rashid,[3] and Nusrat Jahan, 2019, July 11, Major Adverse Cardiovascular Events: An Inevitable Outcome of ST-elevation myocardial infarction, A Literature Review, *Cureus.* 2019 Jul; 11(7): e5280. PMCID: PMC6695291, PMID**:** 31423405
doi: 10.7759/cureus.5280,
Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6695291
21. Kamilu M Karaye, , Mahmoud U Sani, 2008, February 19, Electrocardiographic abnormalities in patients with heart failure, *Cardiovasc J Afr.* 2008 Feb; 19(1): 22–25. PMID: 18320082, PMCID: PMC397531, www.cvjsa.co.za
Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3975312
22. Pleister, Adam & Selemon, Helina & Elton, Shane & Elton, Terry. (2013). Circulating miRNAs: Novel biomarkers of acute coronary syndrome?. Biomarkers in medicine. 7. 287-305. doi:10.2217/bmm.13.8.,

Available at: https://www.researchgate.net/publication/236098928_Circulating_miRNAs_Novel_biomarkers_of_acute_coronary_syndrome
23. Rajalakshmi Santhanakrishnan, Na Wang, Martin G. Larson, Jared W. Magnani, David D. McManus, Steven A. Lubitz, Patrick T. Ellinor, Susan Cheng, Ramachandran S. Vasan, Douglas S. Lee, Thomas J. Wang, Daniel Levy, Emelia J. Benjamin and Jennifer E. Ho, *Atrial fibrillation begets heart failure and vice versa* (no date). Available at: https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.115.018614 (Accessed: April 27, 2023).
24. Soliman, E.Z. *et al.* (2015) "Atrial fibrillation and risk of st-segment–elevation versus non–st-segment–elevation myocardial infarction," *Circulation*, 131(21), pp. 1843–1850. Available at: https://doi.org/10.1161/circulationaha.114.014145.
25. *Atrial fibrillation* (no date) *Mayo Clinic*. Mayo Foundation for Medical Education and Research. Available at: https://www.mayoclinic.org/img-20096412 (Accessed: April 27, 2023).
26. Schnabel, R.B. *et al.* (2010) "Validation of an atrial fibrillation risk algorithm in whites and African Americans," *Archives of Internal Medicine*, 170(21). Available at: https://doi.org/10.1001/archinternmed.2010.434.
27. Hsieh, C.-H. *et al.* (2020) "Detection of atrial fibrillation using 1D convolutional neural network," *Sensors*, 20(7), p. 2136. Available at: https://doi.org/10.3390/s20072136.
28. Liu, X. *et al.* (2021) "Deep learning in ECG diagnosis: A Review," *Knowledge-Based Systems*, 227, p. 107187. Available at: https://doi.org/10.1016/j.knosys.2021.107187.
29. Hochreiter, S. (1998) "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 06(02), pp. 107–116. Available at: https://doi.org/10.1142/s0218488598000094.
30. Petmezas, G. *et al.* (2021) "Automated atrial fibrillation detection using a hybrid CNN-LSTM network on imbalanced ECG datasets," *Biomedical Signal Processing and Control*, 63, p. 102194. Available at: https://doi.org/10.1016/j.bspc.2020.102194.
31. Ryuichiro Yagi, MD, MPH1,2, Shinichi Goto, MD, PhD1,2,3, Yoshinori Katsumata, MD, PhD3 , Calum A MacRae, MD, PhD1,2, Rahul C Deo, MD, PhD 1,2*. Importance of external validation and subgroup analysis in artificial-intelligence detecting low ejection fraction from electrocardiograms. 2022,
Available at: https://www.medrxiv.org/content/10.1101/2022.07.22.22277758v1.full.pdf
32. Morshedi-Meibodi, A. *et al.* (2002) "Heart rate recovery after treadmill exercise testing and risk of cardiovascular disease events (the Framingham Heart Study)," *The American Journal of Cardiology*, 90(8), pp. 848–852. Available at: https://doi.org/10.1016/s0002-9149(02)02706-6.
33. (2022) *UK Biobank - UK Biobank*. Available at: https://www.ukbiobank.ac.uk/ (Accessed: April 27, 2023).
34. *Showcase Homepage*. Available at: https://biobank.ndph.ox.ac.uk/showcase/ (Accessed: April 27, 2023).
35. Zhang Xing Zhong Taiyuan University of Technology *et al.* (2020) *ECG classification using machine learning techniques and smote oversampling technique: Proceedings of the 2020 2nd International Conference on Image Processing and Machine Vision*, *ACM Other conferences*. Available at: https://dl.acm.org/doi/pdf/10.1145/3421558.3421560 (Accessed: April 27, 2023).
36. Nagam, M.R., Vinson, D.R. and Levis, J.T. (2017) *ECG diagnosis: Right ventricular myocardial infarction*, *The Permanente journal*. U.S. National Library of Medicine. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5267627/ (Accessed: April 27, 2023).

37. Yamashita, R. *et al.* (2018) *Convolutional Neural Networks: An overview and application in Radiology*, *Insights into imaging*. U.S. National Library of Medicine. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6108980/ (Accessed: April 27, 2023).
38. Yamashita, R. *et al.* (2018) *Convolutional Neural Networks: An overview and application in radiology - insights into imaging*, *SpringerOpen*. Springer Berlin Heidelberg. Available at: https://insightsimaging.springeropen.com/articles/10.1007/s13244-018-0639-9 (Accessed: April 27, 2023).
39. Phung, V.H. and Rhee, E.J. (2019) *A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets*, *MDPI*. Multidisciplinary Digital Publishing Institute. Available at: https://www.mdpi.com/2076-3417/9/21/4500 (Accessed: April 27, 2023).
40. Alday, E.A.P. *et al.* (2022) *Classification of 12-lead ecgs: The physionet/computing in cardiology challenge 2020*, *Classification of 12-lead ECGs: The PhysioNet/Computing in Cardiology Challenge 2020 v1.0.2*. Available at: https://physionet.org/content/challenge-2020/1.0.2/ (Accessed: April 27, 2023).
41. Duijvenboden, S.van *et al.* (2023) *Prognostic significance of different ventricular ectopic burdens during exercise in asymptomatic UK Biobank subjects*, *medRxiv*. Cold Spring Harbor Laboratory Press. Available at: https://www.medrxiv.org/content/10.1101/2023.03.04.23286713v1 (Accessed: April 27, 2023).
42. Ahmed, A.A. *et al.* (2023) *Classifying cardiac arrhythmia from ECG signal using 1D CNN Deep learning model*, *MDPI*. Multidisciplinary Digital Publishing Institute. Available at: https://www.mdpi.com/2227-7390/11/3/562 (Accessed: April 27, 2023).
43. Nwankpa, C. *et al.* (2018) *Activation functions: Comparison of trends in practice and research for Deep Learning*, *arXiv.org*. Available at: https://arxiv.org/abs/1811.03378 (Accessed: April 27, 2023).
44. *BCELOSS* (no date) *BCELoss - PyTorch 2.0 documentation*. Available at: https://pytorch.org/docs/stable/generated/torch.nn.BCELoss.html (Accessed: April 27, 2023).
45. Kingma, D.P. and Ba, J. (2017) *Adam: A method for stochastic optimization*, *arXiv.org*. Available at: https://arxiv.org/abs/1412.6980 (Accessed: April 27, 2023).
46. *[PDF] the 'K' in K-fold cross validation | semantic scholar* (no date). Available at: https://www.semanticscholar.org/paper/The-'K'-in-K-fold-Cross-Validation-Anguita-Ghelardoni/c19aefe95fc0c120af1f8fad161f640a7f76edae (Accessed: April 27, 2023).
47. Hajian-Tilaki, K. (2013) *Receiver operating characteristic (ROC) curve analysis for Medical Diagnostic Test Evaluation*, *Caspian journal of internal medicine*. U.S. National Library of Medicine. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3755824/ (Accessed: April 27, 2023).
48. Urata, M. *et al.* (2014) *Computed tomography Hounsfield units can predict breast cancer metastasis to axillary lymph nodes - BMC cancer*, *BioMed Central*. BioMed Central. Available at: https://bmccancer.biomedcentral.com/articles/10.1186/1471-2407-14-730/figures/2
49. *ECG-based deep learning and clinical risk factors to predict atrial ...* (no date). Available at: https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.121.057480
50. *Deep neural networks can predict new-onset atrial ... - circulation* (no date). Available at: https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.120.047829
51. Zhu, T. *et al.* (2021) *Minority oversampling for Imbalanced Time Series Classification*, *arXiv.org*. Available at: https://arxiv.org/abs/2004.06373
52. Md Sazzad Hossain *et al.* (2021) *Dual focal loss to address class imbalance in semantic segmentation*, *Neurocomputing*. Elsevier. Available at: https://www.sciencedirect.com/science/article/pii/S0925231221011310.

53. Tianyang Lin and Abstract Transformers have achieved great success in many artificial intelligence fields (2022) *A survey of Transformers*, *AI Open*. Elsevier. Available at: https://www.sciencedirect.com/science/article/pii/S2666651022000146.

Appendix

The code was uploaded and run on Google Colab and/or the UCL Research Computer Services. The code is provided in the supplementary materials. The code is split in section for the data processing part, statistical oversampling and the deep learning algorithms implemented. Certain sections of the code included were only used for development purposes and do not constitute a part of the final code.