Deep Learning Algorithms in the Detection of Cardiovascular Diseases Using Electrocardiogram Data

Shai Aarons
University of Cape Town
Cape Town, South Africa
arnsha011@myuct.ac.za

Jarred Fisher
University of Cape Town
Cape Town, South Africa
fshjar002@myuct.ac.za

Joshua Rosenthal
University of Cape Town
Cape Town, South Africa
rsnjos005@myuct.ac.za

KEYWORDS

Deep Learning, Neural Networks, Cardiovascular Disease

1 PROJECT DESCRIPTION

1.1 Introduction

Statistics published in 2019 by the World Health Organisation indicate that Cardiovascular Diseases (CVDs) account for 32% of global deaths [80]. Approximately 17.9 million people died from CVDs in 2019; of these deaths, 85% were from Myocardial Infarction (MI) or Cerebrovascular accident (CVA) [80]. MI occurs when the blood supply to the myocardium is interrupted, causing extensive damage to the myocardium [44]. The result is often a permanently weakened heart or death if treated too late. Atrial fibrillation (AF) is a common irregular and usually rapid heart rhythm that can lead to coronary blood clots. AF increases the risk of CVA, congestive heart failure, and other heart-related complications [72]. Both AF and MI can be detected by analysing a patient's electrocardiogram (ECG) [61]. An ECG is a real-time, non-invasive technique to measure the heart's electrical activity [30]. ECG monitoring benefits from being implementable with many wearable devices for at-home use [45], making continuous monitoring viable and cost-effective [69].

By detecting CVDs early, patient deterioration can be prevented [82]. Unfortunately, initial misdiagnosis is a significant concern. In one study, almost a third of the 564 412 patients treated for MI were misdiagnosed initially [81]. Moreover, medical facilities may lack expert cardiologists to interpret ECGs [83]. To ease this burden, developing an automated system that can screen patients for CVDs, at a low cost and with relative ease is warranted. The strength of Deep Neural Networks (DNNs) to learn from previous patterns without needing a medical expert to develop the decision logic positions them as effective candidates [23]. DNNs can combine feature extraction and classification into one architecture [63]. This is referred to as the "end-to-end" model (see Figure 1). The end-to-end model is preferred over traditional methods, which require multiple components or expert knowledge [15].

In the last five years, the application of DNNs to CVD detection has increased dramatically and seen substantial successes [71]. However, several limitations persist throughout many such studies: a lack of large, publicly accessible ECG datasets results in many studies training and testing their models on small, singular datasets [61]. The models are thus potentially over-fitted to their dataset, bringing into question their robustness and transference to real-world settings [84]. Additionally, the metrics and validation strategies used in ECG classification are inconsistent [4]. This is an issue as utilising robust evaluation strategies to make results comparable is crucial.

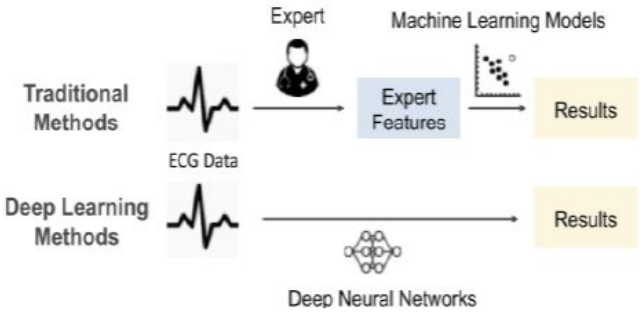

Figure 1: Traditional methods vs. end-to-end models [27]

1.2 Project Overview

Our project endeavours to address the mentioned limitations by creating an experimental platform referred to as *MLECG*. The platform will host an array of DNNs capable of detecting common CVDs, namely MI and AF. The DNNs to be investigated are Recurrent Neural Networks (RNNs) [54] and two of its extensions, i.e. the Long-Short Term Memory (LTSM) [26] and Bidirectional LSTM (BiLSTM) [67]. Additionally, hybrid techniques combining Convolutional Neural Networks (CNNs) [38] and BiLSTMs will also be included in the study.

The relative performances of the models will be compared and stress-tested through various evaluation metrics as described in Section 3.3. The platform will also incorporate a data augmentation scheme capable of correcting for class imbalance in datasets as well as increasing sample sizes. This will be achieved through training a Generative Adversarial Network (GAN) [20] on multiple public ECG datasets. To this end, synthetic ECG data will be made available to the classifiers to augment their training pool.

2 PROBLEM STATEMENT

As mentioned, cases of CVDs are rising globally [9]. Thus there is a need to provide accurate diagnoses of these diseases. Several studies have been done to classify such ailments from ECGs, but to our knowledge, there currently exists no experimental platform that can compare different architecture effectiveness in CVD classification tasks while stress-testing their results. To the best of our knowledge, studies, as seen in the literature, do not attempt to

perform experiments that train models on ECG datasets from one geographical region and test the effectiveness of these models on other ECG datasets from different geographical areas. Furthermore, many previous studies do not consider the prominent issue of class imbalance that is present in ECG data as well as limited dataset size [7]. Experimentation with transfer learning of diseases, geographical regions, or different datasets has seldom been explored in the literature.

2.1 Motivation for Deep Learning Models

2.1.1 Classification Models. The data used as input to RNN models is typically temporal data [37]. Recurrent architectures develop a memory by incorporating past and even future state (in the BiLSTM case [66]) information into the model's current time step [65]. Singh et al. [68] proposed the RNN architecture for Arrhythmia Classification. In contrast to CNNs and other classical ML methods, RNNs can recognise the dependencies between the input and output of a model [64]. RNNs have the advantage that they can identify input signals as time-varying and can therefore capitalise on an ECG's signal's temporal nature as well as its spatial relationships [35]. Thus, an RNN can make inferences on the beat-to-beat variations in addition to the characteristic morphology of the ECG data. As seen in the literature, recurrent models achieved high performance for MI classification and are a sound implementation for detecting CVDs [13, 90].

There have been many variations of hybrid deep learning approaches in the literature, with those joining classical ML with deep learning [8, 86] and those joining two or more deep learning algorithms [2, 52, 56]. The hybrid deep learning architectures, like RNN models, have the advantage of not having feature-engineering modules and also require minimal domain knowledge. It is also evident, from the literature, that the CNN and LSTM combination architecture performs impressively with the task of CVD detection [2, 52, 56]. Ivanovic et al. [32] use a BiLSTM and CNN model to detect AF from private ECG datasets. Their architecture achieved an accuracy of 88.28%. Similarly, Ke Wang et al. [79] used a hybrid CNN and BiLSTM model to detect CVD from ECG data; however, their model was only trained on one dataset, limited to one geographical region. The model achieved an accuracy of 87.69%. There is scope to build these architectures with improved hyperparameters and model structures - perhaps adding additional layers or optimising for efficiency. There is also an opportunity to enhance these models by training on multiple open databases (instead of only one) to provide better and more universal results.

2.1.2 Generative Models. DL techniques require a large amount of fairly distributed training data to succeed [7]. ECG databases are lacking both in terms of size and class balance [46]. Several methods have been used to deal with these issues, such as transfer learning for limited size [19] and random oversampling for class imbalance [76]. Recently, GANs have been used to address both dataset size limitations and class imbalance for DL tasks concerning ECG data to great success [16, 24, 70]. The literature suggests they produce comparatively better performance than other methods [24]. GANs generate data by mapping a random noise vector to a probability distribution [53]. This probability distribution is learned through a minimax zero-sum game between a generator and a discriminator. The generator's goal is to produce samples that deceive the discriminator, and the discriminator's goal is to differentiate between real and fake samples. The game between the two is played until a Nash Equilibrium [51] is met. Where the application of GANs in ECG synthesis can be furthered is through the use of multiple datasets to represent several CVD classes in training, as well as a critical comparison of architectures in the effectiveness of data synthesis.

2.2 Aims & Objectives

This research project aims to investigate a suite of deep learning algorithms by developing a platform that we coin: *MLECG*. The *MLECG* will allow us to perform an array of experiments on the deep learning architectures as mentioned in this proposal. The platform will be modular in that the user will be able to select different parameters for experimentation. These parameters will include the different preprocessing methods used before input to the algorithm and the ability to choose which datasets to train and test the models on, as well as how those datasets will be augmented. The user can also train the model on synthetic data generated by the GAN architectures and test the models on real data or the opposite way around. We aim for the *MLECG* platform to display evaluation metrics of the various models in a concise and readable manner for the experimenter. The goal of the experimental platform is to be able to perform the studies and experiments as discussed in Section 3. These experiments are briefly summarised below:

2.2.1 System.

(1) Use Python 3 [75] to develop, design and implement the MLECG experimental platform

2.2.2 Classification Models.

(1) Perform experiments to discover the optimal hyperparameters needed to achieve the best results in detecting CVDs from ECG data using our described models.
(2) Investigate which preprocessing methods contribute to higher accuracy in the deep learning architectures.
(3) Investigate if increasing the number of hidden layers in the models leads to improved classification abilities for CVDs.
(4) Explore if models trained on ECG data from one geographical region effectively detect CVDs from ECG data from another geographical region.
(5) Use transfer learning to test if training a model with only one class of CVD and then retraining with other CVDs improves overall model performance.

2.2.3 Generative Models.

(1) Ascertain the degree to which WGAN-GP and WGAN architectures stabilise the training process in the context of ECG generation (displaying CVDs) compared to DCGANs under a given hyperparameter search space.
(2) Compare the performance of architectures that use 1D input against those that use 2D input.
(3) Investigate the effectiveness of conditional generation using several datasets for training.

(4) Determine the effectiveness of classifiers trained and tested on synthesised data and real data.

3 PROCEDURE AND METHODS

3.1 Datasets

Without high-quality datasets, deep learning techniques cannot be trained effectively, and they cannot be generalised for use in clinical settings [62]. Concerning supervised learning, samples must be expertly annotated to distinguish healthy control samples from those which are not [85]. This, in addition to the dataset needing to be publicly available, restricts the inclusion of viable datasets for AF and MI detection to the following:

(1) Physikalisch-Technische Bundesanstalt (PTB) Diagnostic ECG Database [5]
(2) MIT-BIH Arrhythmia Database [48]
(3) PTB-XL [78]
(4) Long Term ST Database (LTST) [33]
(5) The China Physiological Signal Challenge 2018 [42]
(6) Telehealth Network of Minas Gerais (15% of original dataset) [60]
(7) The PhysioNet/Computing in Cardiology Challenge 2017 [10]
(8) Huazhong University, Wuhan (only test dataset) [95]

The smallest dataset is the MIT-BIH Arrhythmia Database, consisting of only 48 recordings from 47 patients, [49] whilst the largest database is the Telehealth Network of Minas Gerais, comprising almost two million recordings from 1 676 384 patients [60]. Further details of the eight datasets can be found on Table 1 and Table 2.

Other public ECG datasets used in AF or MI detection were excluded to keep this project's scope to a reasonable level.

3.2 Preprocessing

Since this study aims to work with several different datasets, it is vital that the preprocessing of data is handled carefully. Preprocessing is a crucial step when handling ECG data [36] as it is subject to a range of external artefacts such as baseline wander and power-line interference [41]. To reduce noise, we will look into using filters [50, 94] and discrete wavelet transforms [34, 58, 59]. We will divide ECG recordings into uniform time intervals [17, 24] and, for architectures that require it, transform ECG data into 2D spectrograms [29]. To preserve consistency amongst datasets, we will use normalisation.

3.3 Evaluation of Algorithm Performance

3.3.1 Classification Models. Sensitivity is the most important metric for a diagnostic system to detect a dangerous disease [4]. Using this metric, the system's ability to correctly identify a patient with the disease can be evaluated [30]. The classifier's specificity is closely related because this indicates how effective it is at excluding healthy patients from the diagnosis [84]. There are many other metrics commonly used, which we will implement, such as precision, accuracy, F1 score and area under the receiver operator curve (AUC-ROC) [71].

Best practice in deep learning for classification suggests that any dataset should be partitioned into training, validation, and testing sets with the most amount of data (70%+) being used for the training set [71]. K-fold cross validation is an effective technique to achieve these divisions and is widely used to measure prediction error [18]. We will implement ten-fold cross validation, as this has been used for AF detection with hybrid deep learning models to ensure model generlisability [52] and similarly in MI detection using recurrent architectures [45].

3.3.2 GANs. GANs lack a consistent evaluation metric [14, 16]. This said several measurements have been developed to evaluate the generator that has been used across multiple studies. The methods to be used to directly assess the quality of synthesised ECG data will be Frechét Inception Distance (FID) [25], squared Maximum Mean Distance (MMD) with a Gaussian kernel [21] and Dynamic Time Warping (DTW) [1]. To assess how well synthesised data transfers to classification tasks, the F1 score will be used as a proxy to determine generator efficacy for "train on real, test on synthetic" and "train on synthetic, test on real" experiments [16].

3.4 Classification Models Experiments

The deep learning models to be explored for the classification of CVDs from ECG data are RNN, LSTM, BiLSTM and a hybrid structure combining a CNN module and a BiLSTM module. The classifiers will be tasked with detecting CVDs, namely MI and AF. The datasets to be used can be found in Table 1 and Table 2. The studies will take place over two phases, each with a different purpose. These two phases are discussed in detail below:

3.4.1 Phase One - Architecture and Hyperparameters. Phase one will involve tuning our models to achieve the best possible model performance, in detecting MI and AF, within our given time frame and scope. The performance will be compared using the evaluation metrics as outlined in Section 3.3.1. For this phase, we will use datasets with the least patients such as the MIT-BIH Arrhythmia dataset [49] or the LTST dataset [33] (see Table 1 and Table 2). The smaller datasets will allow for faster training times of the models, allowing for rapid changes to be made to hyperparameters. This will enable us to compare performances using different hyperparameters repeatedly. Firstly, we will segment the ECG data into windows with lengths ranging from 5 seconds to 30 seconds, which is typical of deep learning studies involving ECG data [17]. We will then experiment with different preprocessing methods and which combinations provide the best performance. These preprocessing methods that will be interchanged and combined are outlined in Section 3.2.

Another step to improve algorithm performance is experimenting with different model architecture structures. This will involve assessing if adding additional hidden layers to models improve their performances. We can also experiment with different activation functions such as ReLu or Leaky ReLu. We will additionally experiment with various loss functions such as Cross-Entropy or more novel loss functions such as focal loss, which has proved effective in AF detection in deep learning models [56]. There is also room for experimentation in the convolutional module of the hybrid architecture - including

assessing different pooling methods such as max-pooling or average-pooling. We will similarly evaluate different optimisation strategies, such as the Adam Optimiser, which was used by Ivanovic et al. [32].

We will also experiment with a range of hyperparameters and what changes to these will improve model performance. These hyperparameters include, but are not limited to: initialisation weights, number of epochs and batch sizes.

3.4.2 Phase Two - Datasets and Transfer Learning. The second phase of experimentation will extend our models to different datasets and an assessment of transfer learning. We will use the best performing models found in phase one to perform the following set of experiments:

The first experiment will assess the generalisability of our models to datasets of varying geographical regions. This will determine whether our models are agnostic to geographical location. For this experiment, we will use a dataset from one geographical area for training the models. Then we will use a dataset from a different geographical region to test the models. We will record performance differences and evaluate the generalisability. An example of this methodology would be training the models on the Telehealth Network dataset [60] (the patients in this dataset were local to the Minas region in Brazil) and then testing the models on the Huazhong University dataset [95] (which contains participants local to the Wuhan region in China).

Our second experiment will train the models on a dataset with only one CVD and then use transfer learning to assess whether this could improve the detection of various CVDs. An example would be training the models on the MIT-BIH Arrhythmia dataset [49] (which focuses on Arrhythmia detection) and then retraining the models on the PTB Diagnostic database [5] (which is used widely in MI detection). We will assess if a study setup such as this will improve over phase one's results.

3.5 Generative Model Experiments

The experimentation involving the generative models will consist of two phases.

3.5.1 Phase 1 - Architecture and Hyperparameters. Training of the generative network will be iterative. This will manifest through using increasingly advanced architecture starting from the Deep Convolutional GAN (DCGAN) [57], to the Wasserstein GAN (WGAN) [3], and finally ending with the WGAN using a gradient penalty (WGAN-GP) [22]. All of the mentioned architectures will accept a conditional input [47] so that we might direct the generation process by specifying the class of the signal or image. This will involve embedding diagnostic codes of related ECG samples into the network. In this phase, the models will be trained using the LTST [33] (Table 1). This dataset contains instances of MI and AF and will thus allow us to generate ECGs representing all three desired beat classes namely normal, AF, and MI.

Initially, the architectures used will follow directly from the studies in which they were implemented. This implies that ECG data must be converted into 2D images. This will be accomplished through a process adapted from the study by Brophy et al. [6] using Fourier transforms. Further experimentation on the architecture will involve adjusting the models to accept 1D signals instead of spectrograms [40].

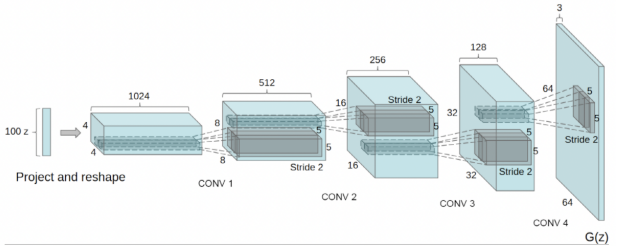

Figure 2: DCGAN generator. [57]

The initial values for hyperparameters will be that of the original DCGAN architecture (Figure 2). The range of hyperparameter adjustments to be made will be based on the work by Lucic et al. [43].

3.5.2 Phase 2 - Datasets. In this phase, we will aim to generate ECGs displaying AF and MI from independent datasets using efficient architectures defined in phase 1. Initial experimentation involving multiple datasets will be adapted from the RadialGAN [89]. More specifically, the discrete case. At a high level, the RadialGAN works by assigning each input domain its own generator and discriminator. It then arranges these constructs around a central noise vector. This particular design choice lends itself well to the iterative design process as datasets can be added onto the radial structure one by one until each domain is represented. The datasets to be represented are displayed in Table 1 and Table 2. Another possible solution to be investigated is to have the models trained using the sample's origin as another label along with the CVD class. For example, an ECG segment displaying AF coming from the MIT-BIH dataset will have both the diagnostic code and the dataset associated with it during training. There is precedent for using multiple conditions to train GANs [16].

3.6 Implementation details

The development of the experimental platform and the architectures of deep learning models will be written in the Python 3 programming language [75]. The deep learning models will be created and coded using the PyTorch library [55]. PyTorch is open source and allows us to create RNNs, hybrid architectures and GANs - thus allowing for greater collaboration and consistency of the different modules being implemented into the experimental platform. We will also use other open source libraries such as SciPy [77] and PyWavelets [39] for the various preprocessing methods discussed above.

4 ETHICAL CONCERNS

The datasets to be used in this research are publicly available through the Open Data Commons Attribution License v1.0 and the Creative Commons Attribution 4.0 International Public License. All datasets contain completely anonymised ECG data. Our research requires no animal or human participants, and no further

data collection is needed. Thus we do not require an ethics clearance. All the software used as a part of this research is free and open source. Programming will be done in Python which is freely available under the Python Software Foundation License. External libraries are given under permissive licenses such as MIT, BSD-like and Apache License 2.0. Where applicable, copyright notices and license agreements will be made accessible in our final codebase.

5 RELATED WORK

5.1 Classification Models

Recurrent architectures develop a memory by combining the output of previous layers with new input to form the current layer's output [54]. The application of recurrent architectures to MI detection is a burgeoning field of study, with the earliest papers only being developed in 2019 [11, 12, 91, 93]. With the limited research available, LSTM based models were identified as the top performer in MI detection, with one study achieving sensitivity, specificity, precision and accuracy values of 99.91% 99.95% 99.91% 99.91%, respectively [91]. The simple RNN design performed the worst in inter and intra study comparisons, achieving the lowest sensitivity, precision and F1 scores of 83%, 68%, and 75%, respectively [11]. Notably, a gap in the literature exists for BiLSTMs and their application to MI detection. Moreover, no study has compared a BiLSTM's performance in MI detection to other architectures, warranting further investigation. A primary concern was that every study that used recurrent architectures for MI detection had trained their models on the public PTB Diagnostic ECG database [5]. This presents an opportunity for research using new ECG datasets.

As mentioned, a hybrid deep learning methodology is when two or more deep learning architectures are combined to, in this context, perform the classification of CVDs. According to Hong et al. [28], the combination of LSTMs and CNNs is the best combination of architectures for detecting AF. Ivanovic et al. [32] proposed a hybrid model incorporating two CNN layers and a BiLSTM layer, achieving an accuracy of 88.28% in detecting AF. Oh et al. [52] demonstrated an impressive accuracy of 98.42% for detecting arrhythmias by using a combined structure of LSTM and CNN layers. While these studies indicate impressive performance, they are either trained on private datasets (such as in Ivanovic et al. [32]) or trained using only one or two datasets. This theme of a lack of diverse datasets is seen through the literature regarding the hybrid method for detecting CVD [2, 32, 52, 56, 87, 88].

Of interest to our research is the application of transfer learning to test whether models trained on one type of CVD classification (such as MI detection) can form the starting point for detecting another CVD such as AF. Significant performance increases have been achieved in CVD classification when using transfer learning [19, 31]. In one study [74] the authors retrained a high-performance image classification model, GoogLeNet [73], on a public ECG dataset, The China Physiological Signal Challenge 2018 [42], and a private dataset of 17 000 ECGs from patients in Southern China with MI [71]. The study attained a reasonable accuracy of 86% on the private dataset but performed poorly on the challenge dataset.

5.2 Generative Models

GANs have been previously shown to be capable of effectively synthesising ECG data. Esteban et al. [16] used GANs to generate medical time series data, achieving MMD statistics as low as 0.35. The authors were able to direct generation using four conditions. Hatamian et al. [24] compared three data augmentation schemes (oversampling, Gaussian Mixture Models, and Deep Convolutional GANs) to be used in DL classification tasks for AF in ECGs. The authors found that classification tasks using data augmented by GANs were the most effective in detecting AF. This study also looked at the effectiveness of CNN-based models when using 1D and 2D input and found those trained using 2D input were more effective. Brophy et al. [6] experimented with a novel method of generating time series data using 2D-based architectures. This process involved transforming ECG signals into 2D images and then using those images to train the generator. Images produced by the generator could then be converted back into signals.

6 ANTICIPATED OUTCOMES

6.1 Research

6.1.1 Classification Models. We anticipate having fully trained and optimised RNN (and its variations) as well as hybrid ML structures, yielding comparable results. We intend to train our models on multiple datasets and optimise preprocessing and model structure to attain the highest possible accuracy and F1 score in our given project time frame. If the timeline is followed, we should have robust architectures capable of detecting CVDs with sufficiently high sensitivity and specificity levels. We also hypothesise that during the second phase of experimentation, our deep learning architectures will be exportable and generalisable across different geographical regions and datasets.

6.1.2 Generative Models. We expect the results of experimentation on generative models to produce ECG data points that transfer well to classification tasks. Furthermore, we anticipate that architectures using the Wasserstein loss function with a gradient penalty in conjunction with 2D input will produce the best results in terms of the evaluation metrics defined in Section 3.3.2 [22, 24]. Finally, we expect the generator to synthesise class-specific data points across several datasets.

6.2 System

We anticipate the *MLECG* to be a modular and robust platform that will allow for a wide array of experiments to be performed on both the generative and classification models. We expect the experimental platform to be thoroughly tested as well as documented.

6.3 Impact

6.3.1 Classification Models. The comparison and assessment of model performance from different geographical regions is a significant gap in the literature. By having a model that transfers across regional and geographical boundaries, we are creating

technology that is exportable across the globe. If our models are successfully trained on multiple large datasets, CVD diagnosis may be performed more effectively. This could relieve some of the burden placed on healthcare systems.

6.3.2 Generative Models. This study partly aims to highlight the potential effect that training stability has on ECG generation by assessing three different architectures. Furthermore, this study aims to tackle the lack of research into the effectiveness of generative models trained over several ECG datasets. As such, we will present a GAN capable of representing CVD-displaying ECGs from a range of datasets.

6.4 Key Success Factors

6.4.1 Classification Models. Key features determining success for the classification models discussed are sensitivity and specificity levels over 90% when tested on ECGs with varying noise and CVD dataset origins. Achieving scores above this threshold would make our models comparable to previous works wherein such successes were found [2, 11, 35, 56, 91, 92]. Another critical success factor is the ability to outline and interpret results from the second phase of transfer-ability experimentation. In terms of design challenges we may face, we anticipate that hyperparameter optimisation and model structure selection will be a time-sensitive process as retraining the model multiple times is a lengthy process. Thus because of this lengthy process, careful consideration of the relevant literature will be imperative in making hyperparameter choices to ensure timely project success.

6.4.2 Generative Models. For the success of the generative models, model convergence is the first and most important factor. Following from this low FID, MMD, and DTW scores will further imply the success of the generator. The generator should be able to produce ECG data representing three discrete classes: Normal, AF, and MI. In conjunction with the system, the success of the generator will be derived from comparable F1 statistics of TSTR and TRTS processes to that of train on real, test on real.

We anticipate several design challenges concerning GANs. Among these challenges is the mode collapse phenomenon, where a generator fails to produce data outside a small subset [16]. This will be counteracted through the use of robust architectures. Furthermore, we expect experimentation with hyperparameters to be complex as GANs are incredibly sensitive [43] and minor changes may result in non-convergence.

7 PROJECT PLAN

7.1 Risk management Strategies

A risk assessment matrix has been outlined in Table 3 to address this.

7.2 Deliverables and Milestones

The key deliverables for this project include the following:

- Three literature reviews
- Project Proposal draft
- Project Proposal final paper
- An initial project feasibility demonstration
- Complete draft of a final paper
- Final project code with written documentation
- Final paper
- Project demonstration
- Project poster
- Project website

There are other significant milestones in our project that include the implementation of the core experimental platform, with its corresponding preprocessing modules, the completion of the coded and optimised deep learning architectures and finally, the completion of the various studies we will be performing.

7.3 Timeline

A Gantt chart, representing our project's timeline with all deliverables and milestones, can be seen in Figure 3.

7.4 Required resources

We will need access to an account for the University of Cape Town High-performance cluster or another high-performance server to train the computationally expensive deep learning models. Otherwise, we will need to leverage Google Collab[1]. However, a free account limits us to a maximum time of 12 hours, and our resource allocation can be throttled. Our last option would be to use a single, high-performance GPU, which we would need to inquire as to the availability of this through UCT or CAIR [2].

7.5 Allocation

All members will collaborate on shared deliverables and milestones, such as the presentation, poster and website creation. Additionally, all members will develop the experimental platform, curate datasets and create the initial preprocessing modules. Thereafter, members will work on their own contributions. Joshua Rosenthal will develop the data augmentation scheme by constructing several GANs. Shai Aarons will focus on hybrid CVD detection techniques, and Jarred Fisher will use recurrent architectures for the same task. Both Shai Aarons and Jarred Fisher will also assess their model's generalisability to geographically separate ECG datasets.

REFERENCES

[1] 2007. *Dynamic Time Warping.* Springer Berlin Heidelberg, Berlin, Heidelberg, 69–84. https://doi.org/10.1007/978-3-540-74048-3_4

[2] Rasmus S. Andersen, Abdolrahman Peimankar, and Sadasivan Puthusserypady. 2019. A deep learning approach for real-time detection of atrial fibrillation. *Expert Systems with Applications* 115 (2019), 465–473. https://doi.org/10.1016/j.eswa.2018.08.011

[3] Martin Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein GAN. https://doi.org/10.48550/ARXIV.1701.07875

[4] Raymond Bond, Dewar Finlay, Salah Shafiq Al-Zaiti, and Peter Macfarlane. 2021. Machine learning with electrocardiograms: A call for guidelines and best practices for 'stress testing'algorithms. *Journal of Electrocardiology* 69 (2021), 1–6.

[5] R. Bousseljot, D. Kreiseler, and A. Schnabel. 1995. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. 40, s1 (1995), 317–318. https://doi.org/doi:10.1515/bmte.1995.40.s1.317

[6] Eoin Brophy, Zhengwei Wang, and Tomas E. Ward. 2019. Quick and Easy Time Series Generation with Established Image-based GANs. https://doi.org/10.48550/ARXIV.1902.05624

[7] Ping Cao, Xinyi Li, Kedong Mao, Fei Lu, Gangmin Ning, Luping Fang, and Qing Pan. 2020. A novel data augmentation method to enhance deep neural networks

[1] https://colab.research.google.com/

[2] https://www.cair.org.za/

for detection of atrial fibrillation. *Biomedical Signal Processing and Control* 56 (2020), 101675. https://doi.org/10.1016/j.bspc.2019.101675

[8] Sucheta Chauhan, Lovekesh Vig, and Shandar Ahmad. 2019. ECG anomaly class identification using LSTM and error profile modeling. *Computers in Biology and Medicine* 109 (2019), 14–21. https://doi.org/10.1016/j.compbiomed.2019.04.009

[9] Sumeet S. Chugh, Gregory A. Roth, Richard F. Gillum, and George A. Mensah. 2014. Global Burden of Atrial Fibrillation in Developed and Developing Nations. *Global Heart* 9, 1 (March 2014), 113. https://doi.org/10.1016/j.gheart.2014.01.004

[10] Gari D Clifford, Chengyu Liu, Benjamin Moody, H Lehman Li-wei, Ikaro Silva, Qiao Li, AE Johnson, and Roger G Mark. 2017. AF classification from a short single lead ECG recording: The PhysioNet/computing in cardiology challenge 2017. In *2017 Computing in Cardiology (CinC)*. IEEE, 1–4.

[11] Annisa Darmawahyuni, Siti Nurmaini, et al. 2019. Deep learning with long short-term memory for enhancement myocardial infarction classification. In *2019 6th International Conference on Instrumentation, Control, and Automation (ICA)*. IEEE, 19–23.

[12] Annisa Darmawahyuni, Siti Nurmaini, Wahyu Caesarendra, Vicko Bhayyu, M Naufal Rachmatullah, et al. 2019. Deep learning with a recurrent network structure in the sequence modeling of imbalanced data for ECG-rhythm classifier. *Algorithms* 12, 6 (2019), 118.

[13] Annisa Darmawahyuni, Siti Nurmaini, Sukemi, Wahyu Caesarendra, Vicko Bhayyu, M Naufal Rachmatullah, and Firdaus. 2019. Deep Learning with a Recurrent Network Structure in the Sequence Modeling of Imbalanced Data for ECG-Rhythm Classifier. *Algorithms* 12, 6 (2019). https://doi.org/10.3390/a12060118

[14] Anne Marie Delaney, Eoin Brophy, and Tomas E. Ward. 2019. Synthesis of Realistic ECG using Generative Adversarial Networks. https://doi.org/10.48550/ARXIV.1909.09150

[15] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. 2015. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2625–2634.

[16] Cristóbal Esteban, Stephanie L. Hyland, and Gunnar Rätsch. 2017. Real-valued (Medical) Time Series Generation with Recurrent Conditional GANs. https://doi.org/10.48550/ARXIV.1706.02633

[17] Xiaomao Fan, Qihang Yao, Yunpeng Cai, Fen Miao, Fangmin Sun, and Ye Li. 2018. Multiscaled Fusion of Deep Convolutional Neural Networks for Screening Atrial Fibrillation From Single Lead Short ECG Recordings. *IEEE Journal of Biomedical and Health Informatics* 22, 6 (2018), 1744–1753. https://doi.org/10.1109/JBHI.2018.2858789

[18] Tadayoshi Fushiki. 2009. Estimation of prediction error by using K-fold cross-validation. *Statistics and Computing* 21, 2 (Oct. 2009), 137–146. https://doi.org/10.1007/s11222-009-9153-8

[19] Ali Ghaffari and Nasimalsadat Madani. 2019. Atrial fibrillation identification based on a deep transfer learning approach. *Biomedical Physics &amp Engineering Express* 5, 3 (mar 2019), 035015. https://doi.org/10.1088/2057-1976/ab1104

[20] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Networks. https://doi.org/10.48550/ARXIV.1406.2661

[21] Arthur Gretton, Karsten Borgwardt, Malte J. Rasch, Bernhard Scholkopf, and Alexander J. Smola. 2008. A Kernel Method for the Two-Sample Problem. https://doi.org/10.48550/ARXIV.0805.2368

[22] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. 2017. Improved Training of Wasserstein GANs. https://doi.org/10.48550/ARXIV.1704.00028

[23] Awni Y Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H Tison, Codie Bourn, Mintu P Turakhia, and Andrew Y Ng. 2019. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine* 25, 1 (2019), 65–69.

[24] Faezeh Nejati Hatamian, Nishant Ravikumar, Sulaiman Vesal, Felix P. Kemeth, Matthias Struck, and Andreas Maier. 2020. The Effect of Data Augmentation on Classification of Atrial Fibrillation in Short Single-Lead ECG Signals Using Deep Neural Networks. https://doi.org/10.48550/ARXIV.2002.02870

[25] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. (2017). https://doi.org/10.48550/ARXIV.1706.08500

[26] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[27] Shenda Hong, Yuxi Zhou, Junyuan Shang, Cao Xiao, and Jimeng Sun. 2020. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Computers in Biology and Medicine* 122 (2020), 103801.

[28] Shenda Hong, Yuxi Zhou, Junyuan Shang, Cao Xiao, and Jimeng Sun. 2020. Opportunities and Challenges of Deep Learning Methods for Electrocardiogram Data: A Systematic Review. https://doi.org/10.48550/ARXIV.2001.01550

[29] Jingshan Huang, Binqiang Chen, Bin Yao, and Wangpeng He. 2019. ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network. *IEEE Access* 7 (2019), 92871–92880. https://doi.org/10.1109/ACCESS.2019.2928017

[30] Pang-Shuo Huang, Yu-Heng Tseng, Chin-Feng Tsai, Jien-Jiun Chen, Shao-Chi Yang, Fu-Chun Chiu, Zheng-Wei Chen, Juey-Jen Hwang, Eric Y Chuang, Yi-Chih Wang, et al. 2022. An Artificial Intelligence-Enabled ECG Algorithm for the Prediction and Localization of Angiography-Proven Coronary Artery Disease. *Biomedicines* 10, 2 (2022), 394.

[31] Ali Isin and Selen Ozdalili. 2017. Cardiac arrhythmia detection using deep learning. *Procedia Computer Science* 120 (2017), 268–275. https://doi.org/10.1016/j.procs.2017.11.238 9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW 2017, 22-23 August 2017, Budapest, Hungary.

[32] Marija D. Ivanovic, Vladimir Atanasoski, Alexei Shvilkin, Ljupco Hadzievski, and Aleksandra Maluckov. 2019. Deep Learning Approach for Highly Specific Atrial Fibrillation and Flutter Detection based on RR Intervals. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 1780–1783. https://doi.org/10.1109/EMBC.2019.8856806

[33] Franc Jager, Alessandro Taddei, George B Moody, Michele Emdin, G Antolič, Roman Dorn, Ales Smrdel, Carlo Marchesi, and Roger G Mark. 2003. Long-term ST database: a reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia. *Medical and Biological Engineering and Computing* 41, 2 (2003), 172–182.

[34] Wissam Jenkal, Rachid Latif, Ahmed Toumanari, A. Dliou, Oussama El B'charri, and Fadel Mrabih Rabou Maoulainine. 2016. An efficient algorithm of ECG signal denoising using the adaptive dual threshold filter and the discrete wavelet transform. *Biocybernetics and Biomedical Engineering* 36 (2016), 499–508.

[35] Debasish Jyotishi and Samarendra Dandapat. 2020. An Attention Based Hierarchical LSTM Model for Detection of Myocardial Infarction. In *2020 IEEE 17th India Council International Conference (INDICON)*. 1–5. https://doi.org/10.1109/INDICON49873.2020.9342488

[36] Michał Kania, Małgorzta Fereniec, and Roman Maniewski. 2007. Wavelet denoising for multi-lead high resolution ECG signals. *Measurement science review* 7, 4 (2007), 30–33.

[37] Gil Keren and Björn Schuller. 2016. Convolutional RNN: An enhanced model for extracting features from sequential data. In *2016 International Joint Conference on Neural Networks (IJCNN)*. 3412–3419. https://doi.org/10.1109/IJCNN.2016.7727636

[38] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436–444.

[39] Gregory Lee, Ralf Gommers, Filip Waselewski, Kai Wohlfahrt, and Aaron O'Leary. 2019. PyWavelets: A Python package for wavelet analysis. *Journal of Open Source Software* 4, 36 (April 2019), 1237. https://doi.org/10.21105/joss.01237

[40] Wenqiang Li, Yuk Ming Tang, Kai Ming Yu, and Suet To. 2022. SLC-GAN: An automated myocardial infarction detection model based on generative adversarial networks and convolutional neural networks with single-lead electrocardiogram synthesis. *Information Sciences* 589 (2022), 738–750. https://doi.org/10.1016/j.ins.2021.12.083

[41] H.-Y. Lin, S.-Y. Liang, Y.-L. Ho, Y.-H. Lin, and H.-P. Ma. 2014. Discrete-wavelet-transform-based noise removal and feature extraction for ECG signals. *IRBM* 35, 6 (2014), 351–361. https://doi.org/10.1016/j.irbm.2014.10.004 Healthcom 2013.

[42] Feifei Liu, Chengyu Liu, Lina Zhao, Xiangyu Zhang, Xiaoling Wu, Xiaoyan Xu, Yulin Liu, Caiyun Ma, Shoushui Wei, Zhiqiang He, et al. 2018. An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. *Journal of Medical Imaging and Health Informatics* 8, 7 (2018), 1368–1373.

[43] Mario Lucic, Karol Kurach, Marcin Michalski, Sylvain Gelly, and Olivier Bousquet. 2017. Are GANs Created Equal? A Large-Scale Study. https://doi.org/10.48550/ARXIV.1711.10337

[44] Waldomiro Carlos Manfroi, Carolina Peukert, Clarissa Bacha Berti, Clarissa Noer, Danielle de Ávila Gutierres, and Felipe Theodoro Bezerra Gaspar Carvalho da Silva. 2002. Acute myocardial infarction: the first manifestation of ischemic heart disease and relation to risk factors. *Arquivos Brasileiros de Cardiologia* 78, 4 (2002), 392–395.

[45] Harold Martin, Walter Izquierdo, Mercedes Cabrerizo, Anastasio Cabrera, and Malek Adjouadi. 2021. Near real-time single-beat myocardial infarction detection from single-lead electrocardiogram using Long Short-Term Memory Neural Network. *Biomedical Signal Processing and Control* 68 (2021), 102683.

[46] Barbara Mukami Maweu, Rittika Shamsuddin, Sagnik Dakshit, and Balakrishnan Prabhakaran. 2021. Generating Healthcare Time Series Data for Improving Diagnostic Accuracy of Deep Neural Networks. *IEEE Transactions on Instrumentation and Measurement* 70 (2021), 1–15. https://doi.org/10.1109/TIM.2021.3077049

[47] Mehdi Mirza and Simon Osindero. 2014. Conditional Generative Adversarial Nets. https://doi.org/10.48550/ARXIV.1411.1784

[48] George B Moody and Roger G Mark. 1992. MIT-BIH Arrhythmia Database. https://doi.org/10.13026/C2F305

[49] George B Moody and Roger G Mark. 2001. The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine* 20, 3 (2001), 45–50.

[50] Deepankar Nankani and Rashmi Dutta Baruah. 2020. Investigating Deep Convolution Conditional GANs for Electrocardiogram Generation. In *2020 International Joint Conference on Neural Networks (IJCNN)*. 1–8. https://doi.org/10.1109/IJCNN48605.2020.9207613

[51] J.F. Nash. 1951. Non-cooperative Games. *Annals of Mathematics* 54, 2 (1951), 286–295.

[52] Shu Lih Oh, Eddie Y.K. Ng, Ru San Tan, and U. Rajendra Acharya. 2018. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Computers in Biology and Medicine* 102 (2018), 278–287. https://doi.org/10.1016/j.compbiomed.2018.06.002

[53] Zhaoqing Pan, Weijie Yu, Bosi Wang, Haoran Xie, Victor S. Sheng, Jianjun Lei, and Sam Kwong. 2020. Loss Functions of Generative Adversarial Networks (GANs): Opportunities and Challenges. *IEEE Transactions on Emerging Topics in Computational Intelligence* 4, 4 (2020), 500–522. https://doi.org/10.1109/TETCI.2020.2991774

[54] Razvan Pascanu, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. 2013. How to construct deep recurrent neural networks. *arXiv preprint arXiv:1312.6026* (2013).

[55] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 8024–8035. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[56] Georgios Petmezas, Kostas Haris, Leandros Stefanopoulos, Vassilis Kilintzis, Andreas Tzavelis, John A Rogers, Aggelos K Katsaggelos, and Nicos Maglaveras. 2021. Automated Atrial Fibrillation Detection using a Hybrid CNN-LSTM Network on Imbalanced ECG Datasets. *Biomedical Signal Processing and Control* 63 (2021), 102194. https://doi.org/10.1016/j.bspc.2020.102194

[57] Alec Radford, Luke Metz, and Soumith Chintala. 2015. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. https://doi.org/10.48550/ARXIV.1511.06434

[58] Hari Mohan Rai and Anurag Trivedi. 2012. De-noising of ECG Waveforms based on Multi-resolution Wavelet Transform. *International Journal of Computer Applications* 45 (2012), 25–30.

[59] Hari Mohan Rai, Anurag Trivedi, and Shailja Shukla. 2013. ECG signal processing for abnormalities detection using multi-resolution wavelet transform and Artificial Neural Network classifier. *Measurement* 46, 9 (2013), 3238–3246. https://doi.org/10.1016/j.measurement.2013.05.021

[60] Antônio H. Ribeiro, Gabriela M.M. Paixao, Emilly M. Lima, Manoel Horta Ribeiro, Marcelo M. Pinto Filho, Paulo R. Gomes, Derick M. Oliveira, Wagner Meira Jr, Thömas B Schon, and Antonio Luiz P. Ribeiro. 2021. CODE-15%: a large scale annotated dataset of 12-lead ECGs. (June 2021). https://doi.org/10.5281/zenodo.4916206

[61] Beanbonyka Rim, Nak-Jun Sung, Sedong Min, and Min Hong. 2020. Deep learning in physiological signal data: A survey. *Sensors* 20, 4 (2020), 969.

[62] Beanbonyka Rim, Nak-Jun Sung, Sedong Min, and Min Hong. 2020. Deep Learning in Physiological Signal Data: A Survey. *Sensors* 20, 4 (2020). https://doi.org/10.3390/s20040969

[63] K Sanjana, V Sowmya, EA Gopalakrishnan, and KP Soman. 2020. Explainable artificial intelligence for heart rate variability in ECG signal. *Healthcare Technology Letters* 7, 6 (2020), 146.

[64] Jürgen Schmidhuber. 2015. Deep learning in neural networks: An overview. *Neural Networks* 61 (jan 2015), 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

[65] Robin M. Schmidt. 2019. Recurrent Neural Networks (RNNs): A gentle Introduction and Overview. https://doi.org/10.48550/ARXIV.1912.05911

[66] M. Schuster and K.K. Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing* 45, 11 (1997), 2673–2681. https://doi.org/10.1109/78.650093

[67] Mike Schuster and Kuldip K Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing* 45, 11 (1997), 2673–2681.

[68] Shraddha Singh, Saroj Kumar Pandey, Urja Pawar, and Rekh Ram Janghel. 2018. Classification of ECG Arrhythmia using Recurrent Neural Networks. *Procedia Computer Science* 132 (2018), 1290–1297. https://doi.org/10.1016/j.procs.2018.05.045

[69] Konstantinos C Siontis, Peter A Noseworthy, Zachi I Attia, and Paul A Friedman. 2021. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nature Reviews Cardiology* 18, 7 (2021), 465–478.

[70] Kaleb E Smith and Anthony O Smith. 2020. Conditional GAN for timeseries generation. https://doi.org/10.48550/ARXIV.2006.16477

[71] Sulaiman Somani, Adam J Russak, Felix Richter, Shan Zhao, Akhil Vaid, Fayzan Chaudhry, Jessica K De Freitas, Nidhi Naik, Riccardo Miotto, Girish N Nadkarni, et al. 2021. Deep learning and the electrocardiogram: review of the current state-of-the-art. *EP Europace* 23, 8 (2021), 1179–1191.

[72] Mayo Clinic Staff. 2021. *Atrial Fibrillation*. Retrieved May 16, 2022 from https://www.mayoclinic.org/diseases-conditions/atrial-fibrillation/symptoms-causes/syc-20350624#:~:text=Atrial%20fibrillation%20(A%2Dfib),and%20other%20heart%2Drelated%20complications.

[73] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9.

[74] Girmaw Abebe Tadesse, Tingting Zhu, Yong Liu, Yingling Zhou, Jiyan Chen, Maoyi Tian, and David Clifton. 2019. Cardiovascular disease diagnosis using cross-domain transfer learning. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 4262–4265.

[75] Guido Van Rossum and Fred L. Drake. 2009. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.

[76] Dhwaj Verma and Sonali Agarwal. 2018. Cardiac Arrhythmia Detection from Single-lead ECG using CNN and LSTM assisted by Oversampling. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. 14–17. https://doi.org/10.1109/ICACCI.2018.8554541

[77] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17 (2020), 261–272. https://doi.org/10.1038/s41592-019-0686-2

[78] Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Dieter Kreiseler, Fatima I Lunze, Wojciech Samek, and Tobias Schaeffter. 2020. PTB-XL, a large publicly available electrocardiography dataset. *Scientific data* 7, 1 (2020), 1–15.

[79] Eric Ke Wang, liu Xi, Ruipei Sun, Fan Wang, Leyun Pan, Caixia Cheng, Antonia Dimitrakopoulou-Srauss, Nie Zhe, and Yueping Li. 2019. A new deep learning model for assisted diagnosis on electrocardiogram. *Mathematical Biosciences and Engineering* 16, 4 (2019), 2481–2491. https://doi.org/10.3934/mbe.2019124

[80] WHO. 2019. *Global Health Estimates: Life expectancy and leading causes of death and disability*. Retrieved May 17, 2022 from https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates

[81] Jianhua Wu, Chris P Gale, Marlous Hall, Tatendashe B Dondo, Elizabeth Metcalfe, Ged Oliver, Phil D Batin, Harry Hemingway, Adam Timmis, and Robert M West. 2018. Editor's Choice-Impact of initial hospital diagnosis on mortality for acute myocardial infarction: A national cohort study. *European Heart Journal: Acute Cardiovascular Care* 7, 2 (2018), 139–148.

[82] Lin Wu, Guifang Huang, Xianguan Yu, Minzhong Ye, Lu Liu, Yesheng Ling, Xiangyu Liu, Dinghui Liu, Bin Zhou, Yong Liu, et al. 2022. Deep Learning Networks Accurately Detect ST-Segment Elevation Myocardial Infarction and Culprit Vessel. *Frontiers in cardiovascular medicine* 9 (2022).

[83] Ran Xiao, Yuan Xu, Michele M Pelter, David W Mortara, and Xiao Hu. 2018. A deep learning approach to examine ischemic ST changes in ambulatory ECG recordings. *AMIA Summits on Translational Science Proceedings* 2018 (2018), 256.

[84] Ping Xiong, Simon Ming-Yuen Lee, and Ging Chan. 2022. Deep Learning for Detecting and Locating Myocardial Infarction by Electrocardiogram: A Literature Review. *Frontiers in Cardiovascular Medicine* 9 (2022).

[85] Yun Xu and Royston Goodacre. 2018. On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of analysis and testing* 2, 3 (2018), 249–262.

[86] Weiyi Yang, Yujuan Si, Di Wang, and Gong Zhang. 2019. A Novel Approach for Multi-Lead ECG Classification Using DL-CCANet and TL-CCANet. *Sensors* 19, 14 (2019). https://doi.org/10.3390/s19143214

[87] Weiyi Yang, Yujuan Si, Di Wang, and Gong Zhang. 2019. A Novel Approach for Multi-Lead ECG Classification Using DL-CCANet and TL-CCANet. *Sensors* 19, 14 (2019). https://doi.org/10.3390/s19143214

[88] Ozal Yildirim, Ulas Baran Baloglu, Ru-San Tan, Edward J. Ciaccio, and U. Rajendra Acharya. 2019. A new approach for arrhythmia classification using deep coded features and LSTM networks. *Computer Methods and Programs in Biomedicine* 176 (2019), 121–133. https://doi.org/10.1016/j.cmpb.2019.05.004

[89] Jinsung Yoon, James Jordon, and Mihaela van der Schaar. 2018. RadialGAN: Leveraging multiple datasets to improve target-specific predictive models using Generative Adversarial Networks. https://doi.org/10.48550/ARXIV.1802.06403

[90] Xingjin Zhang, Runchuan Li, Qingyan Hu, Bing Zhou, and Zongmin Wang. 2019. A New Automatic Approach to Distinguish Myocardial Infarction Based on LSTM. In *2019 8th International Symposium on Next Generation Electronics (ISNE)*. 1–3. https://doi.org/10.1109/ISNE.2019.8896550

[91] Xingjin Zhang, Runchuan Li, Qingyan Hu, Bing Zhou, and Zongmin Wang. 2019. A new automatic approach to distinguish myocardial infarction based on LSTM. In *2019 8th International Symposium on Next Generation Electronics (ISNE)*. IEEE,

1–3.

[92] Xingjin Zhang, Runchuan Li, Qingyan Hu, Bing Zhou, and Zongmin Wang. 2019. A New Automatic Approach to Distinguish Myocardial Infarction Based on LSTM. In *2019 8th International Symposium on Next Generation Electronics (ISNE)*. 1–3. https://doi.org/10.1109/ISNE.2019.8896550

[93] Yue Zhang and Jie Li. 2019. Application of Heartbeat-Attention Mechanism for Detection of Myocardial Infarction Using 12-Lead ECG Records. *Applied Sciences* 9, 16 (2019), 3328.

[94] Zhi-dong Zhao and Yu-quan Chen. 2006. A New Method for Removal of Baseline Wander and Power Line Interference in ECG Signals. In *2006 International Conference on Machine Learning and Cybernetics*. 4342–4347. https://doi.org/10.1109/ICMLC.2006.259082

[95] Hongling Zhu, Cheng Cheng, Hang Yin, Xingyi Li, Ping Zuo, Jia Ding, Fan Lin, Jingyi Wang, Beitong Zhou, Yonge Li, Shouxing Hu, Yulong Xiong, Binran Wang, Guohua Wan, Xiaoyun Yang, and Ye Yuan. 2020. Automatic multilabel electrocardiogram diagnosis of heart rhythm or conduction abnormalities with deep learning: a cohort study. *The Lancet Digital Health* 2, 7 (2020), e348–e357. https://doi.org/10.1016/S2589-7500(20)30107-2

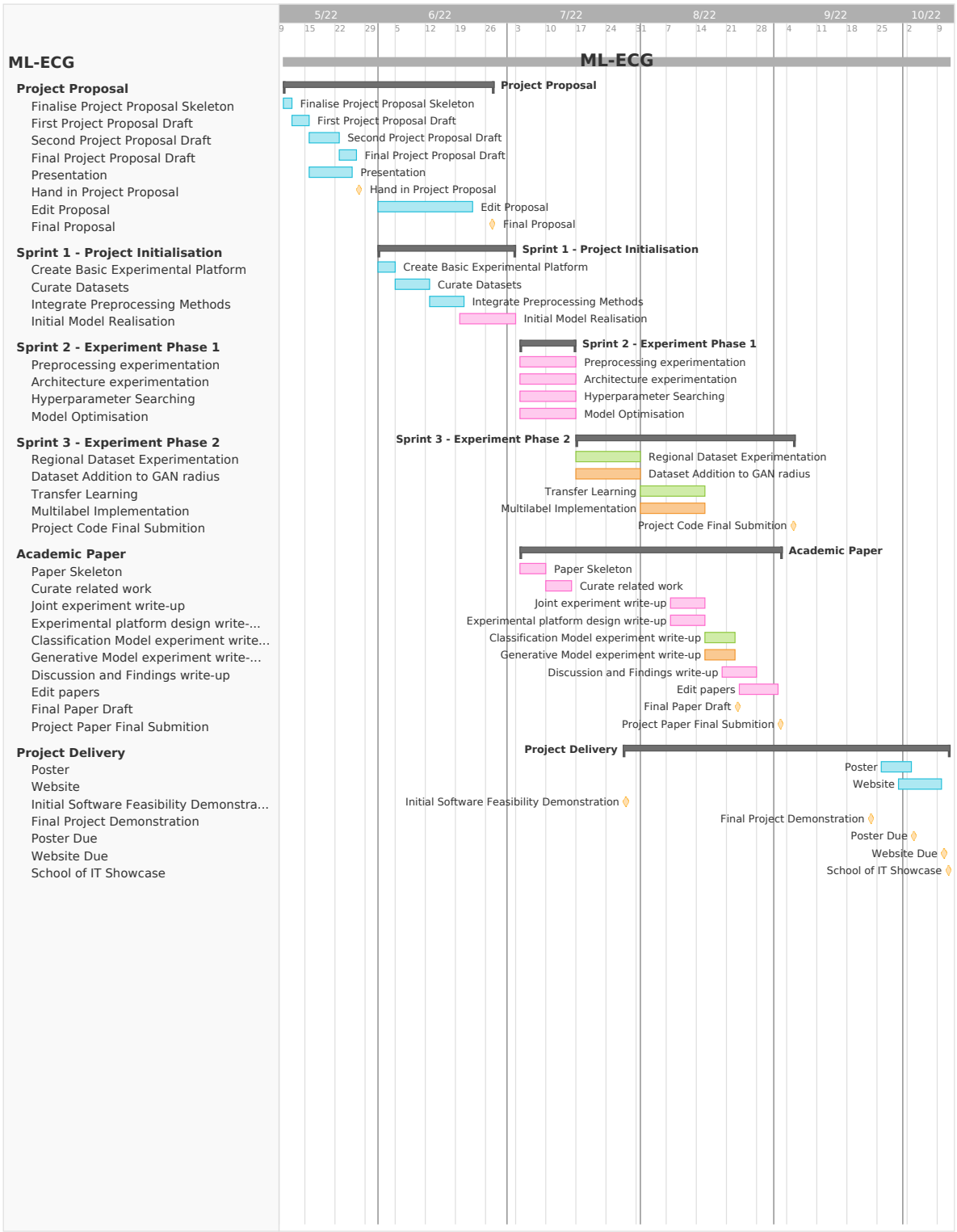

Figure 3: Gannt Chart representing project timeline.

Table 1: Public ECG datasets

Dataset	Study period	Country	Application to our research	ECG Information	Patient Count	Record Count	Strengths	Limitations
LTST [33]	2003-2007	Slovenia	MI and AF detection	Leads: 2 or 3 Length: 21-24 hours Frequency: 250Hz Resolution: 12 bits with +-10 mV range Annotations: locations of the PQ junction (the isoelectric level) and the J point, ST level time series or the ST deviation time series	**80**	**86**	- Detailed clinical notes and ST deviation trend plots are provided for all 86 records. - Contains both AF and MI labelled data. Allows testing of model's capability to detect AF and MI within the same cohort.	- Only 2 or 3 leads used -Data sample size is small. - MI and healthy control not explicitly labelled.
MIT-BIH Arrhythmia [48]	1975-1979	USA	AF detection	Leads: 2 Patient Demographics: Aged 23 - 89; 25 men, 22 women. Length: 30 minutes Frequency: 360Hz Resolution: 12 bits with +-10 mV range. Annotations: QRS complex, R-R intervals and peaks, PQ junction and J point, signal quality, beat, diagnosis and rhythm classifications.	**47**	**48**	- Extensive annotations - Expert cardiologists independently annotated each record.	- Small sample size. - Only 2 leads used
The China Physiological Signal Challenge 2018 [42]	2018	China	AF detection	Leads: 12 Patient Demographics: 3699 male, 3178 female recordings Length: 6 – 60 seconds Frequency: 500Hz Annotations: QRS complex, R-R intervals and peaks, PQ junction and J point, signal quality, beat, diagnosis and rhythm classifications.	-	**6877**	- Large sample size and well annotated. - ECG data comes from 11 different hospitals – more generalisable as hospitals may use various ECG equipment, thus introducing real world noise and variability.	- Test set excluded. - Minimal ECG information given: Missing resolution, patient and age information.
Telehealth Network of Minas Gerais (15% stratified sample of original dataset) [60]	2010-2016	Brazil	AF detection	Leads: 12 Patient Demographics: Mean age: 51.6; 40.2% male, 59.8% female. Length: 7 – 10 seconds Frequency: 400 Hz Resolution: Unknown Annotations: Beat and diagnosis classifications.	**233770**	**345779**	- Extremely large sample size. - High quality dataset.	- Labelling of ECG abnormalities was done with imperfect automatic coders and natural language processing models. – Not all labels have been reviewed by cardiologists. - No P-QRS-T signal annotations given

Table 2: Public ECG datasets cont.

Dataset	Study period	Country	Application to our research	ECG Information	Patient Count	Record Count	Strengths	Limitations
PTB [5]	2000	Germany	MI detection	Leads: 12 leads + 3 Frank leads Patient demographics: Aged 17–87, mean 57.2; 209 men, mean age 55.5, 81 women, mean age 61.6 Length: 2 minutes Frequency: 1 kHz-10 kHz Resolution: 16 bits with 0,5 V/LSB (2,000 A/D units per mV) Annotations: Rhythm and diagnosis classification.	**290**	**549**	- 15 leads ECG used. - MI and healthy control patients are explicitly labelled. - Higher resolution than LTST - Noise levels were recorded during signal collection	- Small sample size. - Data is only recorded from a single site. - No P-QRS-T signal annotations were given.
The PhysioNet/ Computing in Cardiology Challenge 2017 [10]	2017	USA	AF detection	Leads: Single lead Patient demographics: Unknown Length: 9 – 60 seconds Frequency: 300 Hz Resolution: 16-bit files with a bandwidth of 0.5-40 Hz and a ± 5 mV dynamic range. Annotations: Rhythm and diagnosis classification.	-	**8528**	- Large sample size - High resolution - Single lead ECGs can be found in everyday fitness devices. A model achieving high sensitivity on this dataset would be readily deployable on wearable tech.	- Test set excluded. - Minimal ECG information is given. Missing resolution and patient demographic information. - No P-QRS-T signal annotations were given. - Only one cardiologist oversaw labelling. Mid challenge, the trustworthiness of the labelling was questioned and later revealed discrepancies.
PTB-XL [78]	1989-1996	Switzerland	MI detection	Leads: 12 Patient demographics: Ages: from 0 to 95 years Median 62 and IQR of 22 Male: Female ratio = 52:48% Length: 10 seconds Frequency: 500Hz Metadata: demographics, infarction features, likelihoods for diagnostic ECG statements and annotated signal properties	**18885**	**21837**	- Largest open access 12-lead ECG waveform dataset. - Recorded at multiple sites. - MI and healthy control records are explicitly labelled. - Suggests folds for splitting training and testing data. This incentivises standardisation for evaluating model performance.	
Tongji Hospital: Huazhong University, Wuhan (Test dataset only) [95]	2012-2019	China	AF detection	Leads: 12 Patient demographics: Mean age 50.8; Male: Female ratio = 41:59% Length: 10 seconds or 24 hours Frequency: 500Hz Resolution: Unknown Metadata: Rhythm and diagnosis classification.	**828**	**828**	- Three board-certified actively practising cardiologists, including one cardiac electrophysiologist, annotated ECGs in this test dataset. - Extensive labels representing different ECG rhythm classes. - Vast difference in recording time amongst samples allows for an intra-dataset comparison of classification ability when trained on short vs. long term ECG data.	- Small sample size on account of only having public access to test dataset. - Minimal ECG information is given. Missing resolution and age information. - No P-QRS-T signal annotations are given.

Table 3: Risk matrix

Risk	Probability	Impact	Consequence	Mitigation	Monitoring	Management
Inability to complete tests and validation	Low	Catastrophic	Unable to draw sound conclusions. Findings will be severely limited.	Ensure deadlines are adhered as specified by the project time line.	Pay careful attention to project development status in relation to project timeline.	Prioritise completion of pivotal project components.
Project partners withdraw from project	Low	Medium	Shared workload would have to be redistributed. This would have cascading effects on the proposed timeline.	Limit task dependencies between members.	Regular meetings with project partners to assess headspace and progress.	Communicate with supervisors and reduce project scope.
Insufficient time to adequately train models	Low	High	Unable to obtain comparable results to previous findings.	Use high-quality hardware to reduce overall training time.	Compare expected training times with initial training times.	Plan accordingly using more accurate estimates of training times.
Lack of engagement with supervisor.	Low	Medium	Lack of guidance may lead the project astray - possibly leading to low scientific contribution	Maintain regular communication with supervisors. Schedule regular meetings.	Observe the rate of scheduled meeting attendance.	Schedule regular meetings with supervisors every 2 weeks.
Course work interferes with project progress	Medium	Medium	Delays in project deliverables - pushing time line forward. Fewer scientific conclusions ultimately drawn.	Careful planning and time management to allow for concurrency between development and course work.	Pay attention to course work deadlines and their interference with project deliverables.	Allowing for slack time in project development cycle.
Load Shedding slowing development progress	High	Low	Unable to train models or work on project paper further delays time line.	Plan training intervals in accordance with load shedding schedules.	Use the Eskom publically released load shedding schedules.	Work at UCT during load shedding hours as a generator is provided.
Project member gets Covid-19	Low	Low	Individual contributions suffer. Further delays in time line as a result.	Limit task dependencies between members. Practice recommended hygiene routines.	Pay careful attention to physical well-being.	Slack time permits for suboptimal human resources.
Scope creep	High	Medium	Time delays result in lower productivity performing meaningful tasks.	Careful planning and communication with supervisors.	Verify that any proposed changes to project design fall in line with the overarching objectives.	Prioritise completion of pivotal experiments and features. Drop extraneous research.