Imperial College London

Imperial College London

School of Public Health

Deriving AF phenotypes clusters from AI-ECG methods on patients with persistent atrial fbrillation

Author: cid(02364467)
Word count: 9356

Submitted in partial fulfilment of the requirements for the MSc degree in Health data analytics and machine learning Imperial College London

September 2023

1. Acknowledgements

I would like to thank the members of the ElectroCardioMaths Programme department at Imperial College London for their help throughout the project. The consistent meetings and positive research environment were essential in pushing me into new fields of research. I would also thank them for the novel set of data they have collected previously there were only 14 patients' worth of data to work on now there are 285.

Abstract

Background: As the burden of Cardiovascular disease continues to increase late-stage persistent atrial fibrillation has become more of a focus. The current gold standard for treating persistent atrial fibrillation is catheter ablation which provides informative intracardiac signals but only boosts success rates of 53.6% at most. The inability to successfully treat longstanding and persistent AF has led to a focus on attempting to discover novel approaches to dealing with this disease. One looming issue is the lack of categorisation for a disease that expresses as much heterogeneity as AF. A solution to this is AF phenotypes, the prevailing idea is to cluster patient information to develop new subgroups of AF that can better inform treatment. Therefore, the primary objective is to evaluate methods of predicting the morphology of an intracardiac lead (CS3-4) from 4 ECG leads (I,aVF, V1,V6) using ECG-AI models and generate a set of features from those predictions. This contributes to the secondary objective of extracting features from these predictions and analysing the characteristics of these clusters to evaluate their clinical relevance.

Methods: 62,100, 2-second samples were used to derive a prediction using a CNN for the sample entropy of the intracardiac lead based on 4 ECG leads. The features were then extracted for 49,530 of those samples that had valid corresponding patient info. K means clustering was used on these features to provide 7 clusters and a series of statistical tests(Wald rank sum test,chi-squared test and Fischer exact test) were used to identify if the distribution of these clusters where significantly different to the first cluster which was used as a reference cluster

Results: For both the autoencoder solution and the CNN approach the predictive performance was weak however the CNN managed to predict sample entropies well for sample entropies lower than 0.5 but overall out of 1 the correlation for each test set was 0.276 on average. We have identified 4 statistically significant clusters. Clusters 3 and 6 had lower means for AF risk factors and comorbidities. Clusters 5 and 7 showed a higher means of risk factors like atrial diameter being highest at 1.95 cm for cluster 5.

Discussion: Overall we were able to establish phenotypes with differing levels of risk factors and comorbidities of AF. These can be used to aid in more specific research questions to these 4 phenotypes like gearing research of a drug towards a specific specific AF phenotype These results open up the potential of identifying these phenotypes in a non-invasive approach however the sample entropy prediction performance must improve before we can do this. Improvements in model methodology such as using different loss functions like the Pearson correlation coefficient might help us to do so.

Keywords: Atrial Fibrillation, Catheter Ablation, ECG, CNN, Autoencoder

2. Introduction

Atrial fibrillation is one of the most common forms of cardiac arrhythmia leading to an increased risk of stroke event by 3 – 5 folds even when adjusting for external risk factors. (Wolf et al., 1978). Based on statistics from 2017, the prevalence of the disease is at 37,574 million and the incidence is now 31% higher than the incidence rate in 1997 at 403/million inhabitants of incidence (Lippi, Sanchis-Gomar and Cervellin, 2021). There are different types of atrial fibrillation (AF) dependent on the amount of time a patient has been dealing with the disease. Paroxysmal is categorized by a short-term (under a week) return to normal sinus rhythm. Persistent atrial fibrillation, a focus point of this research, can last more than 7 days or more than a year which it is then considered to be longstanding persistent AF. Lastly, permanent AF is usually considered when there seems to be no hope of the heart returning to normal sinus rhythm as decided by a health professional. The majority of the prevalence is made up of people with permanent AF as seen in this European study which has found that persistent AF is prevalent in 25% of people while

permanent AF occurs in 50% of people with AF (Zoni-Berisso et al., 2014). The current gold standard for treating cases of atrial fibrillation that struggle to respond to anti-arrhythmic drugs is Catheter ablation, this is especially successful with people with paroxysmal AF. The catheter ablation procedure requires placing a catheter into the heart through the groin and performing conventional pulmonary vein isolation in order to ablate the tissue responsible for causing an irregular rhythm. This has shown a great amount of success in preventing people from progressing to persistent AF, in which a systematic review has found that people who did not undertake the procedure had a between 10 – 20% chance of progressing after 1-year follow-up while people who undertook the procedure had a 2.4 – 2.7 % chance of progression over 5 years of follow up (Proietti et al., 2015). These promising results are not reflected when it comes to catheter ablation procedures for people with persistent AF, with trials like the CAPLA trial only showing an effective treatment (no atrial arrhythmia lasting for more than 30 seconds) for 53.6 per cent of patients that had undergone catheter ablation and no change when paired with antiarrhythmic drugs (Chieng et al., 2022). It is reasonable to suggest that the majority prevalence of permanent AF is partially due to an inability to effectively treat patients with persistent AF which has driven research into finding more effective methods of treatment or looking at ways to better diagnose a patient in order to suggest a pathway of current treatment earlier on in the hopes of increasing the success rate of a procedure like catheter ablation. Throughout this paper, we will further explore improving current methods and attempt to add to the current knowledge pool.

2.1 A deeper exploration of atrial fibrillation

To understand this thesis it is important to understand the basics, this entails looking over what an ECG is, what AF is , how it presents itself in a patient and what is catheter ablation.

2.1.1 What is an ECG?

The standard practice for taking an ECG involves using 10 cables/electrodes in order to obtain a 12-lead surface ECG. The leads labeled v1-v6 are placed on the chest and cover the areas closest to the heart seen in Figure 1. The 4 other electrodes are all placed on the extremities and drive leads with the labels: I, II,

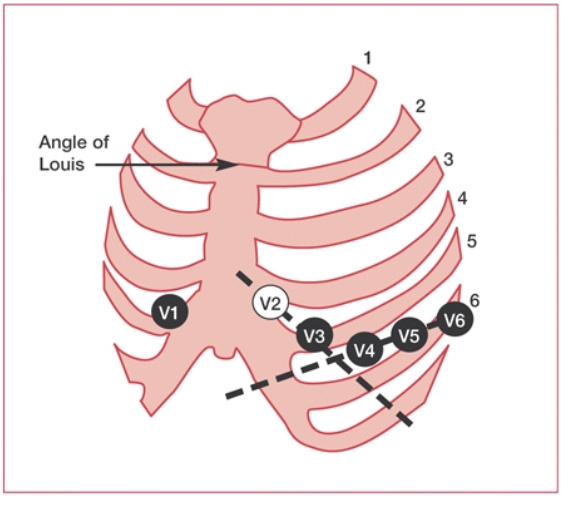

Figure 1. This figure shows the placement of the 6 electrodes that generate the 6 leads: V1-6. This diagram has been taken from the third chapter of the Cardiology Explained book Conquering ECGs(Ashley and Niebauer, 2004).

III, aVR, aVL, and aVF . Three of these are placed on the distal limbs to measure a potential difference between limbs and the final electrode acts as a ground electrode placed on the spare right ankle to reduce

background noise when recording the ECG. The limb leads I – III measure a potential difference between the limbs for example lead I measure the potential difference between the right and left arm, and the augmented leads (aVR,aVL,aVF) are calculated by finding the potential difference of one of the three limbs to an estimate of zero potential. This creates a representation of the vertical plane of the heart as seen in the figure below.(Ashley and Niebauer, 2004)

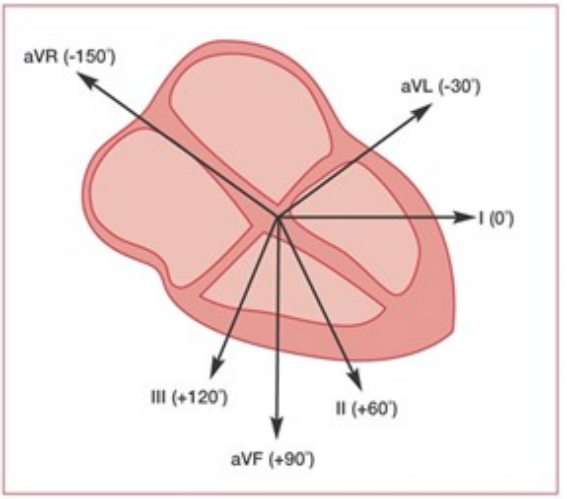

Figure 2. This figure represents the vertical plane of the heart. The arrows correspond to one of the leads generated from the electrodes on the distal limbs showing what lead represents what part of the heart.This diagram has been taken from the third chapter of the Cardiology Explained book Conquering ECGs(Ashley and Niebauer, 2004).

2.1.2 What does an ECG reading look like and what does it mean

Figure 3 below gives you a general idea of what an ECG should look like. The P wave acts to represent atrial depolarization which then can lead to depolarization of the ventricles seen in the QRS segment. We expect and hope the QRS will be narrow to represent an efficient depolarization of the ventricles. A wider QRS may suggest less efficient depolarization due to dysfunction in the conduction system. ('ECG interpretation: Characteristics of the normal ECG (Rashwani,2023). During AF the atria are fibrillating at exceedingly fast rates from 400 to 600 beats per minute. The speed of this means the threshold for the activation of the ventricles to fill up and contract is not always met reducing ventricular activation and subsequently presenting as a reduced amount of QRS waves relative to P waves.

2.1.3 What is pulmonary vein isolation ablation

Pulmonary vein isolation has become the current gold standard in treating difficult-to-cure AF and as previously mentioned has shown a high level of success with people with paroxysmal AF. Pulmonary veins are located on the left side of the heart they are responsible for taking in oxygen-rich blood from the lungs into the heart, the entrance point for these veins into the heart is called Ostia. The majority of Humans normally will have 4 pulmonary veins however this number can differ between 2-6(Klimek-Piotrowska et al., 2016). Pulmonary veins are effective targets of ablation as they have been linked to the pathogenesis of AF in multiple ways.Not all mechanisms of AF have been explored but there is some current research avalaible. One proposed issue is re-entry into the Pulmonary veins, this has been linked to alteration in the electrophysiological properties of the muscle cells responsible for Pulmonary vein blood entry into the heart(Mahida et al., 2015). Surgeons will use the intracardiac reading from an EGM forming an EAM (3d electro-anatomical map) to dictate where to ablate.

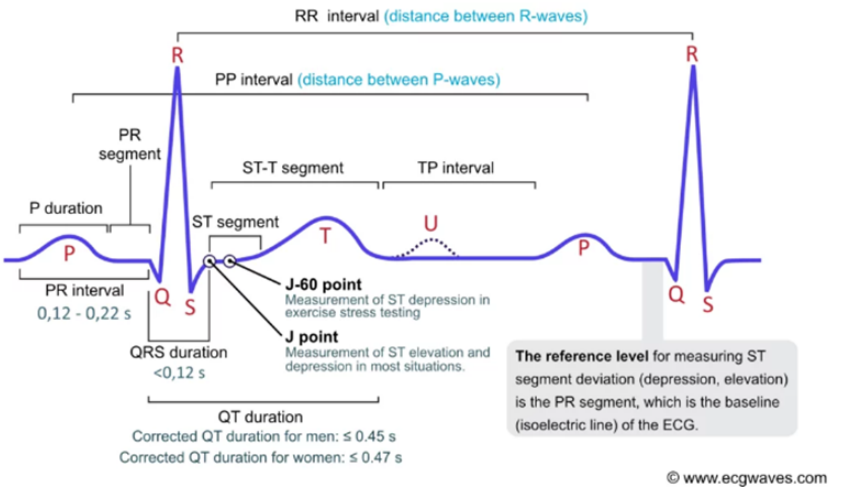

Figure 3. This figure shows the structure of a typical healthy ECG signal each structure is clearly labeled. The start of the p wave towards the end of the T wave is considered one cardiac cycle.This figure is supplied by the clinical ECG interpretation book Section 1 Chapter 5 on the ECG and ECHO learning website (Rashwani,2023)

There is always an exploratory stage to treatment for people with AF from the drug treatment to the ablation methodology. We have seen this in ablation in which a secondary ablation is required due to failure of the first ablation(Escribano et al., 2022). It would be more effective if we had more specificity in diagnosing subcategories of AF so that we can identify the most likely treatment pathway and rely less on an exploratory approach. In the next section, we will introduce current concepts and research behind creating new subcategories for AF commonly referred to as AF phenotypes.

2.1.4 The idea behind AF phenotypes

Researchers have made attempts to find different ways to create additional phenotypes for AF based on certain characteristics of patients, the main solution for this is clustering. This method involves using machine learning to cluster data identifying groups of patients with similar characteristics. There has been some clustering to identify new AF phenotypes, one study has found 4 clusters which had been determined by CV(cardiovascular) risk factors and comorbidities. Cluster 1 had low rates of both however 2 and 3 were characterised by the high burden of CV risk factors and comorbidities with cluster 3 having a much higher number of comorbidities than cluster 2 and cluster 4 was defined by a high level of non-CV comorbidities which contained an older group of individuals seen in the cluster. From this cluster analysis, we can determine that in a treatment situation cluster 2 , given its lower number of comorbidities but higher amount of risk factors would benefit from early treatment and change of lifestyle factors whereas cluster 4 who are already suffering from many comorbidities and where not responding to drug treatment which may need to go forth and look for ablation as a solution (Vitolo et al., 2021). Another study identified clusters like the first study in which the first cluster contained young men with a low prevalence of CV comorbidities however this study identified 5 clusters as opposed to 4 with gender also being a deciding factor in addition to cardiovascular risk factors and comorbidities(Saito et al., 2023). From looking at just two studies that share similarities and differences between their clusters we can see that AF is not a simple disease and can present itself in different ways among different patients and therefore requires tailored treatment for each type of patient. We will later explore how we can build on this research by creating clusters of our own but to understand this we must first cover the fundamentals of deep learning and how we will be utilizing them within this thesis.

2.2 Deep learning and ECGs

As researchers sometimes we are handed non-linear data due to the complex patterns in nature. As such, we must develop new statistical methods to deal with these non-linearities hence the great focus of research on deep learning.

2.2.1 What is deep learning?

These networks are comprised of an input layer and a hidden layer that adds to the true depth within the model and is comprised of multiple neurons and multiple layers that serve the purpose of detecting and processing the different patterns within the data. Finally, there is an output layer that provides the output required, sometimes that can be a classification prediction but it can also be a linear prediction of something like forecasting what the stock market may look like in the coming days. Like humans machines rarely get it right the first time and so these models are paired up with an optimizer function, these functions require a loss metric that can reflect the accuracy of a prediction which allows the system to adjust the parameters of the network accordingly by adjusting the weights associated with the neurons in the hopes that these new weights can help the model reach a more accurate output(Alzubaidi et al., 2021). This problem is referred to as the gradient problem which reflects how much the parameters must change (whether decreasing or increasing) to minimize the cost/loss function. Each neuron has a local gradient calculated by the chain rule which gives each neuron within every layer a local gradient of the partial derivatives of the outputs relative to its input. Finally to optimize the local gradients an algorithm often called gradient descent, adjusts the weights and biases of neurons which are initialized randomly but then are iteratively changed with the sole purpose of minimizing the cost function (Kostadinov, 2019). This algorithm also uses the learning rate, another pre-determined user-defined number that decides how much influence the gradient has at each iteration, this number is normally kept small, especially with large complex problems as to keep the model generalizable (perform well on untrained data) to a broad range of patterns across a large dataset(Wilson and Martinez, 2001). This number of iterations through the data are called epochs. Methods like early stopping will be used to make sure the model is optimized sufficiently. Usually, model users will have training data to train the model, a small segment of data as validation and a final section to really test the effectiveness of the model called the test dataset. In early stopping the model is trained on the training data but is stopped when the loss function of validation data stops improving over a certain number of epochs this will make sure that the model is also generalizable by preventing the model from overfitting to the training data set(Bai et al., 2021).

2.2.2 What is a CNN

The benefit of deep neural networks is that they can be modified for more specialised tasks. CNNs are convolutional neural networks that contain convolutional layers within them. These layers are able to capture a things like spatial relationships in the data which in the case of ECGs is a temporal relationship. Usually, the convolutional layers are 2D and therefore are specialised for computer vision tasks (photo analysis). However, when we deal with ECGs, especially when they are digitally recorded, we then are dealing with a 1D array of values with the length of the array corresponding to the time. Previously, if you had a 1D array, it was essential to transform the data into two dimensions and then to fit 2D Convolutional layers which were inefficient. 1D convolutional networks serve the purpose of requiring no such preprocessing you can just put your raw 1D signal in and have features automatically extracted for the purpose of things like classification. Instead of 2D matrices that are created by 2D Conv layers 1D Conv layers produce an output of 1D convolutional sequence that represents a weighted sum of two 1D arrays. Not dealing with 2D matrices allows us to have a more efficient backpropagation process and therefore have an efficient way of extracting and analysing features from a raw 1D signal like an ECG (Kiranyaz et al., 2019). Using 1D CNN has shown a lot of promise in detecting atrial fibrillation, there have been CNN models that have been tested on MIT-BIH atrial fibrillation database and have achieved

over 97% in sensitivity and specificity when classifying signals into normal. Atrial fibrillation, atrial flutter and AV junctional rhythm(Petmezas et al., 2021).

2.2.3 What is an autoencoder?

An autoencoder still utilizes the CNN architecture but in a different way. The model architecture consists of an encoder, a decoder and in the middle is a latent space. The encoder, aided by convolutional layers can act to dilute the input into lower dimensionality which makes up the latent space. Essentially it is a dimensionality reduction into a latent space However, with this method, unlike principal component analysis (PCA) , we can use non-linear transformation (due to using CNN architecture) to capture more complex and intricate relationships between the data as we reduce to a smaller set of dimensions. The decoder then works to reconstruct this signal and the loss function (mean absolute error) is based on the model's ability to reconstruct the signal as accurately as possible in which backpropagation is also used the adjust the network parameter to provide the most accurate reconstructed signal that is possible(Bank, Koenigstein and Giryes, 2021). The promise of this model is that it can provide a more intricate set of features that can be clustered on or used for classification purposes. Researchers have shown promising success in classifying types of arrhythmias using feature extraction, in which researchers have achieved 97% accuracy in classifying 6 versions of arrhythmia including : normal sinus beat, atrial fibrillation, ventricular bigeminy, pacing beat, atrial flutter and sinus bradycardia(Ramkumar et al., 2022).

2.3 What is missing and how can we add to it (research aims)?

The primary objective of this thesis will be to generate features relevant to the reconstruction of the intracardiac rhythm from 4 lead ECGs. This will lead to our secondary objective which is to evaluate clustering methods upon these features and to analyse these clusters to derive new phenotypes related to risk factors, comorbidities and drug history. This thesis explores two methods to achieve the primary objective of intracardiac reconstruction. The first method required reconstructing the intracardiac rhythm from an EGM reading using a catheter reading taken during ablation in unison with a 4-lead ECG. This work is building on the work by Banta et al who have been able to reconstruct cardiac cycles for a 12-lead ECG from five (also 1) EGM leads and vice versa using autoencoders(Banta et al., 2021). We attempt to reconstruct the full intracardiac signal from the full ECG signal as oppose to just reconstructing the cardiac cycle using autoencoders. The advantage over the alternative method, which we will mention next, is being able to reconstruct a patient's intracardiac recording from an ECG will provide us with a non-invasive method for viewing an intracardiac reading. These intracardiac reading can help surgeons establish a location in which to ablate so being able to reconstruct this recording non-invasively can mitigate the exploratory process during ablation improving the speed at which it is done(Koulouris and Cascella, 2023). The alternative method will involve calculating sample entropy which is a representation of the complexity of a time series signal, this has previously been used on simulated intracardiac signals to detect functional reentry during AF(Ugarte, Tobón and Orozco-Duque, 2019). In this paper, we will explore how we can use a CNN model to predict the intracardiac sample entropy using the corresponding 4 lead ECG signals. Both these methods serve the purpose of providing features in some way for the secondary objective of cluster analysis and phenotype derivation. Being able to identify clusters using either of these methods will help healthcare professionals identify the AF phenotype of that patient in a non-invasive way based on their predicted intracardiac features. We have mentioned before how clustering has been able to suggest new phenotypes based on comorbidities and risk factors for AF. Here we will be able to cluster using features related directly to the intracardiac rhythm (signal or sample entropy) potentially giving us a non-invasive method to determine AF phenotypes based solely on an ECG and directly linked to the intracardiac rhythm of each patient. In this, we looked at deep learning-based approaches to see how clustering can help us derive phenogroups based on the characteristics of someone's surface ECG reading relative to their intracardiac readings. We did this to see if we can put patients into clinically relevant phenogroups. Being able to establish relevant groups based on the ECG may help us

decide the treatment strategy that may be needed for future patients based on their ECG alone.

3. Methods

3.1 Data

The data was collected from 285 patients by the ElectroCardioMaths department at Imperial College London. All the patients used had persistent or longstanding AF and all went through pulmonary vein isolation procedures. Information from the procedure has been recorded including the method used for ablation (RF or cryoablation) , was the patient in AF after the ablation and was there a recurrence of Arrhythmia after the procedure had been completed. The ECG and EGM were stored in a TXT and Python(V3.10.8) was used to extract the ECG data. Initially, before pre-processing, 2415 samples were extracted from 285 unique patients.

3.2 Study focus and reasoning

In this study, we focus on four leads as the procedure in which data has been collected resulted in the 4 leads being more well aligned with the intracardiac reading while the 12-lead ECG was not aligned. The four leads chosen were: I, aVF, V1 and V6. Both I and V6 represent the lateral surface of the heart while V1 gives signals related to the right atrium and cavity of the left ventricle and aVF is related to the inferior surface of the heart(Meek and Morris, 2002). . Having been informed by the surface ECG leads, surgeons can now begin the invasive exploratory phase which is the beginning of catheter ablation where the catheters are now inserted into the groin and are used to obtain an intracardiac reading from the coronary sinus (CS). At the same time, a 4-lead surface ECG is also being recorded which is what makes up the samples of our dataset containing 4 leads ECG and corresponding EGM intracardiac lead. The intracardiac lead we look at is CS 3-4 which is located near the middle of the coronary sinus (Di Marco et al., 2013). CS 3-4 was also the lead that was most available lead being present in 98% of 2,415 60-second samples available for 285 patients.

3.3 Study pipeline

The flow chart seen in Figure 5 has detailed the study pipeline from the initial set of 2415 60-second samples the the final set of 62,100 2-second samples.

Figure 4. This figure breaks down the study flow of this thesis. First, the patents required some pre-processing which involved removing some patients based on exclusion criteria such as not all 60,000 timestamps being available, and requiring the presence of the 4 leads and CS lead that is required which resulted in 2,100 samples for 285 unique patients. Then the samples were split into 2 seconds each so that there were 30 2-second samples per 60-second signal which finally resulted in 62,100 samples for 285 unique patients. The unique patients were split into 5 and each 1/5th of the patients was put in a fold which resulted in making sure every patient was assigned one fold to avoid any data leakage. This resulted in a slight mismatch in total samples in each fold as you can see in the figure however the imbalance was not large enough to cause any concern as the goal was to prevent any data leakage (having samples from one patient in more than one fold) was more of a priority to ensure a robust methodology. Then the folds were organised in sets where the test fold would have the penultimate fold, unless it was fold 1, as the validation data and the rest of the folds were used for training the model. For both the autoencoder and CNN regression the training a validation sets were used to train the model and the validation set was used to ensure some generalizability of the model through early stopping in which the performance was assessed on the training set and validation data. A model was trained for each set 1-5 to produce a prediction for the test set of the corresponding fold essentially producing a prediction for each fold as the test set in which the performance of the output was assessed also. Due to the lack of performance seen in the autoencoder, feature extraction was only done for the regression CNN, in which 512 features were extracted from the test set predictions of each model. The samples for patients that had patient information on them and that met the quality control measures were used which brought the dataset down to 196 unique patients and 49530 samples. The features from the samples obtained from the test set prediction were then clustered and a cluster analysis was completed to identify characteristics for each cluster. More details of the clustering process can be found under the clustering subsection in the methods section.

3.4 Deep learning model methodology

Part of the primary objective of this study was to reconstruct or predict the morphology of the intracardiac rhythm either the whole signal or the sample entropy. To do this two models were used the autoencoder had the goal of reconstructing the signal completely and the CNN had the goal of predicting sample entropy. Both these models were used with the end goal of extracting features from these models for clustering. This section will detail the thinking behind the model structures.

3.4.1 Autoencoder

An autoencoder model was adapted from the work of Kuznetsov et al who used a variational autoencoder to generate a new cardiac cycle (QRS peak and small proportion surrounding signal including the p and t wave) that are similar to the original cardiac cycles the model was trained on(Kuznetsov et al., 2021). However, as we only want to reconstruct the exact ECG signal or intracardiac rhythm that we want, the variational (sampling) component of the model was removed as we did not need to generate new signals we just needed to recreate them. Due to the larger signal size of 2000, we used larger kernel sizes resulting in a larger latent space. The architecture was developed in TensorFlow (v2.7.1) in collaboration with CUDA (V11.4.1) in order to utilise an Nvidia graphics card(RTX6000). To create the architecture we used Oleszaks work who shows what a basic autoencoder should look like, we modified this architecture to use more layers to output a linear activation function and use a 1dimensional input rather than a two 2-dimensional input(Oleszak, 2023). The details of the model will be depicted below. Within the encoder,

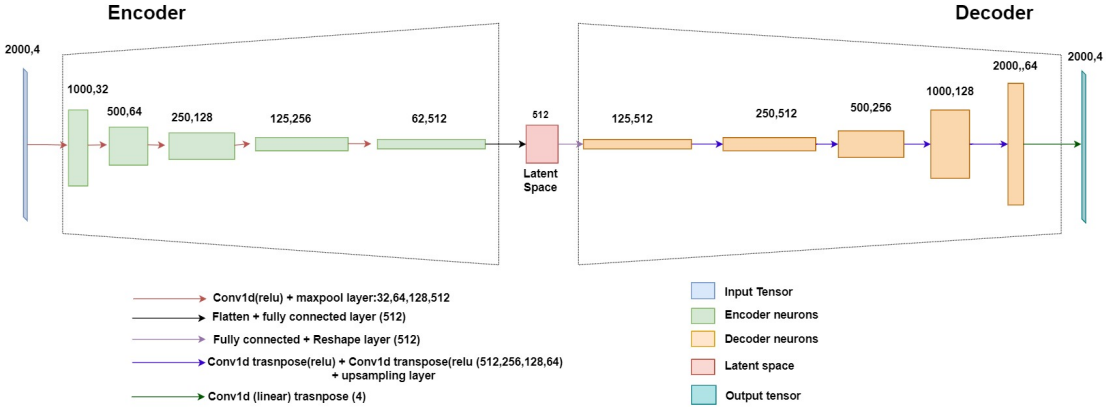

Figure 5. This figure displays an autoencoder model that details the autoencoder segment into the latent space the the decoder segment into the output tensor. The boxes within the encoder, and decoder represent the shape of the output tensor after the layer specified in the arrow legend is applied to the tensor at each stage. The number next to the arrow layer specification represents the kernel size used in the order they were used in and within the brackets next to the specified layer will be the activation function used e.g Conv1d(activation function) 32 kernel size of the first array,64 kernel size of the second arrow. The max pool layer does not have a kernel size so the numbers next to it correspond to the Conv1D layer, not the max pool layers

a convolutional 1D layer was used . While I have detailed the backpropagation in the introduction, it is useful to explain the forward propagation to explain how features are extracted. Covolutional 1D layers use a 1d kernel that is passed across a 1D array to create a 1D feature map(Kiranyaz et al., 2021). The Maxpool layer had a pool size of 2 so for 2 features the max number is taken and the other is discarded, halving the dimensions of the output e.g from 1000 to 500. After this, a flatten and dense(fully connected) layer is used to flatten the output to 512 as this represents the latent space of 512 of the most important features extracted from the signal. For the decoding, a dense layer with a size of 512 is used to expand the latent space to 64000 which is reshaped back to 125,512. Then gets put through a transposed convolution 1d layer which essentially reverses the convolutional process and up-samples the data so that we can up-sample

the feature map back into the required output(Lane, 2018). Finally, a Conv1D transpose layer of size 4 with a linear activation function, producing an output of any number (negative or positive) is used to upsample to the original 4 lead ECG. In the case of the reconstruction of the intracardiac rhythm the same process is undertaken however the transpose layer is reduced to a size of 1 resulting in an output of 2000,1 to represent one 2000 timestamped EGM signal.

3.4.2 CNN

The alternative approach to the primary objective of feature extraction is to use the ECGs to predict the sample entropy of the Intracardiac reading rather than to reconstruct it. To do this we use a standard CNN architecture to predict the sample entropy from an input of 4 ecg leads. This architecture is adapted from (Khan et al., 2023) who leverage a resnet architecture to classify different types of heartbeat from an ECG at an f1 score of 92.83 which shows improvements over other CNNs. We determined that it would be helpful to utilize this model as it seemed amicable in detecting morphological areas of interest within an ECG in the hopes that it would be able to identify different morphologies in our 4 lead ECG and relate that to the sample entropy. The model is shown below.

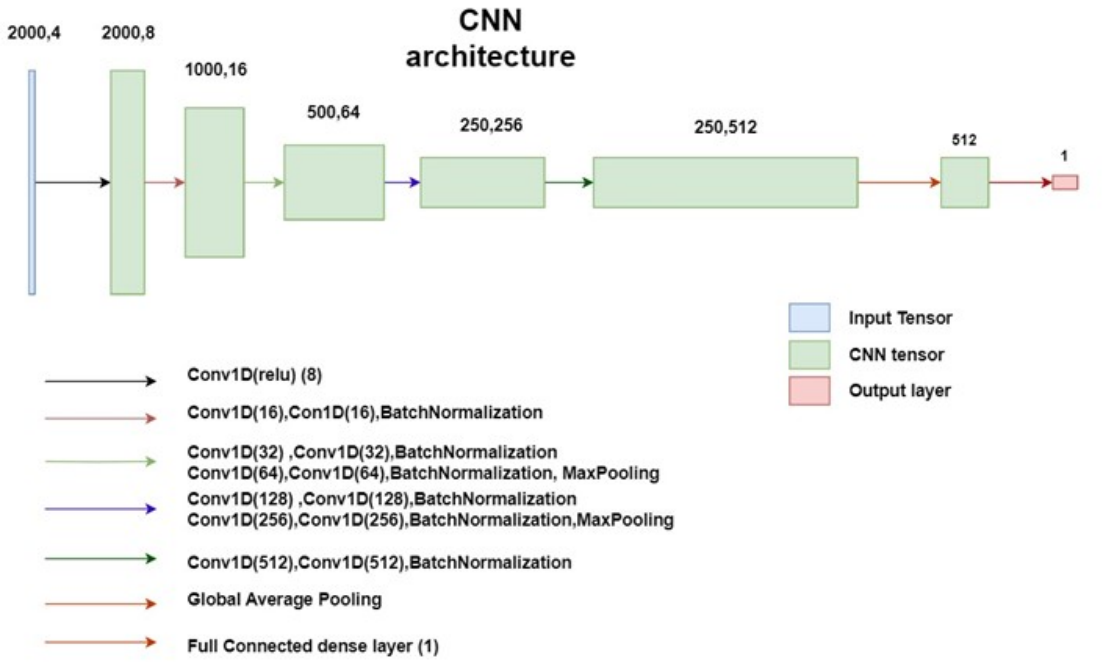

Figure 6. This figure represents the CNN architecture used to predict sample entropy. The arrows represent the layers applied to the tensor and in brackets is the kernel size per layer e.g Conv1D (kernel size)

3.5 Feature extraction, clustering and cluster analysis

3.5.1 Feature Extraction and clustering

Our secondary objective is to extract significant features and identify new af phenotypes through clustering upon those features. The features are extracted from the penultimate layer of the CNN as seen in the study pipeline each fold (1-5) will be used to extract features based on the test set prediction. We then cluster through a series of clustering techniques:Kmeans,GMMC and DBscan in which based on performance (silhouette score) we picked Kmeans. We used the elbow method to identify the most stable cluster number

which is 8 however we did go for 7 clusters as adding more clusters will make some clusters way too small with only 2 samples which would cause too much imbalance between the clusters.

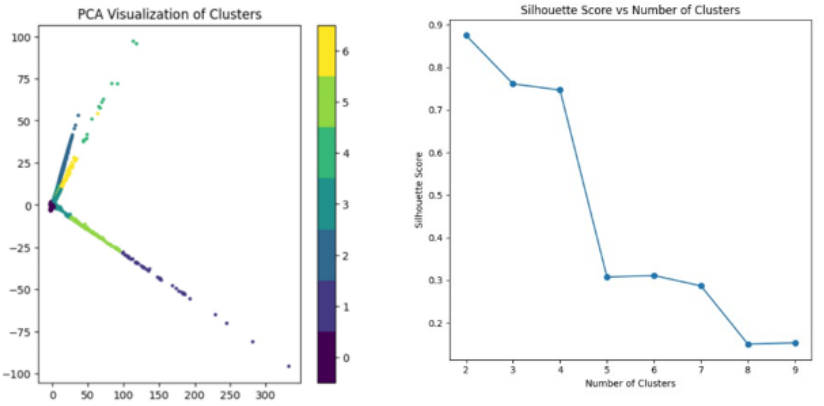

Figure 7. This figure depicts the cluster performance of the feature extraction. The left plot shows a dimensionally reduced PCA visualisation of the clusters and the right plot represents the silhouette score per cluster. Kmeans is a very common clustering method that gained prominence in many fields during the 1950s-60s. Kmeans cluster works by initially going by a user-defined amount of cluster (k). We then need to minimise the sum of the square error of the cluster itself to do this the K-means algorithm picks a random Euclidian space to place (k) number of centroids. Each point of data is assigned to the nearest Euclidean distance centroid and the algorithm iteratively picks out these centroid locations until there are stable clusters formed that are maximally separated from each other(Ikotun et al., 2023). The silhouette score is then used as a graph to assess the compactness and degree of separation between clusters (Rousseeuw, 1987).

3.5.2 Data preparation

A csv was supplied by the electrocardiomaths department for 196 patients resulting in features from 49,530 being analysed. The imputation process used here was miss Forest due to its adept ability to deal with different data types whether they were continuous or categorical in which it has performed better that other imputation algorithms like KNNimputation(Stekhoven and Bühlmann, 2012). Normally the threshold for imputing missingness is 30 % but we needed to impute some important variables that could not be removed like LA volume which has a higher percentage of missingness above 45% so we used the out of bag error to determine if the imputation was acceptable. The out-of-bag error for numerical variables is represented by the normalized root mean squared error which essentially measures the numerical variance between the real data set and the imputed data set and can be measured between 0 and 1 here it was 0.1685 which is deemed acceptable. The categorical column imputation performance was the proportion of falsely classified (PFC) at 0.0098 which was also small and deemed acceptable for this data (Stekhoven, 2022).

3.5.3 Cluster analysis

Each set of features per sample was set to 1 of the seven clusters. Repeated samples belonging to one patient would be repeated leading to the patient's repeated data being added to the cluster again. Cluster 1 was considered the reference cluster as each patient had been assigned this cluster based on at least 10 samples from that patient, so it served as a baseline for the most common features among patients and the other clusters represent what distribution of features differentiate a set of samples from the baseline distribution. For numerical variables, the statistical test used was the Wilcox rank sum test which is used to compare two samples from each other. This is considered a non-parametric test as we do not assume the samples are normally distributed(Xia, 2020). In this test, the samples are combined and ranked in which the sum of ranks is calculated for each sample and a test statistic (U) is used based on the lowest sum of ranks of

between the samples. An expected value of the test statistic is calculated and variance between the real and expected is used to calculate the Z-score which is used to derive the P value. If the P value is under 0.007 then it is statistically significant and we can reject the null hypothesis that two values do have the same continuous distribution(Wilcoxon Test: Definition in Statistics, Types, and Calculation, 2023). For categorical values the Pearson chi-squared test is used and for clusters that are smaller like cluster 2,5,3 and 6 the Fischer exact test is used. The Pearson chi-squared and Fischer test uses the calculated expected and observed values if there is a large significant difference between the observed and expected values then the distribution of the categorical values are considered significant from each other.

4. Results

In this section, we will evaluate the performance of our models for both the autoencoder-based methods and the CNN-based methods. Looking through the results we will provide suitable reasoning for our methodologies through this paper and provide insights into what the results show when it comes to the cluster analysis.

4.1 Evaluation of autoencoder performance on ECG reconstruction

All 62,100 2 second sample were used to evaluate the ECG performance this will provide some interpretability and understanding of what features the autoencoder keeps and what feature it is ignoring from the ECG . This will help us evaluate if the autoencoder is sufficiently picking up the ECG signal.

We can see that the performance when it comes to the training and validation loss curve of the models overall is good for this task of reconstruction. The large improvements in validation and training loss occur within the first 20 epochs and there are then small incremental improvements after that(figure8). The models last for a relatively large number of epochs lasting to about more than 100 epochs setting up a guideline for how many epochs may be required for reconstruction being 200 or less(figure8). As for the distribution of MAEs, we can see that in general, the models produce reconstruction mostly between the 0-200 range, this is seen in the total average MAE measurements and the individual lead reconstruction performance(figure 9).It is important to note some outliers larger than and MAE of 500 with the individual lead performance of V1 and V6 showing an MAE of up to 5000 however the proportion of outliers in relatively small compared to the whole dataset(figure 9). Overall the performance of the ECG reconstruction based on the MAE performance and the train validation loss is adequate and gives us a good idea of the quality of the reconstruction however it is worth looking at some of the reconstruction to further understand the autoencoder performance Looking at both we can see that the reconstruction for an ECG that is clean with no noise performs very well(figure 10). There seems to be a noise reduction effect on the ECG the small bumps that represent a minuscule increase in the signal between the peaks are removed and replaced by a smoothed outline in the reconstruction. This difference is evident in the actual avF and the reconstructed avF for the good reconstruction (figure 10). For the badly performing reconstruction when those small changes are slightly larger as seen in the original aVF lead, the model struggles to reconstruct the signal however it is able to reconstruct the QRS peaks sufficiently(figure 11). We also see the effect of noise on the reconstruction as the original V6 is improper and considered to be noise and this leads to a bad V6 reconstruction(figure 11). It is of note the many signals above 400 MAE are dealing with noise in one of the leads, normally V1 and V6 , hence producing reconstruction responsible for those outliers seen in the distribution plots for the MAEs of the reconstructions(figure 9). So even with noise the reconstruction performs very well visually. It is worrying that the autoencoder smoothes out the signal between the peaks as this may be essential in informing the structure of the intracardiac rhythm beyond the peaks.

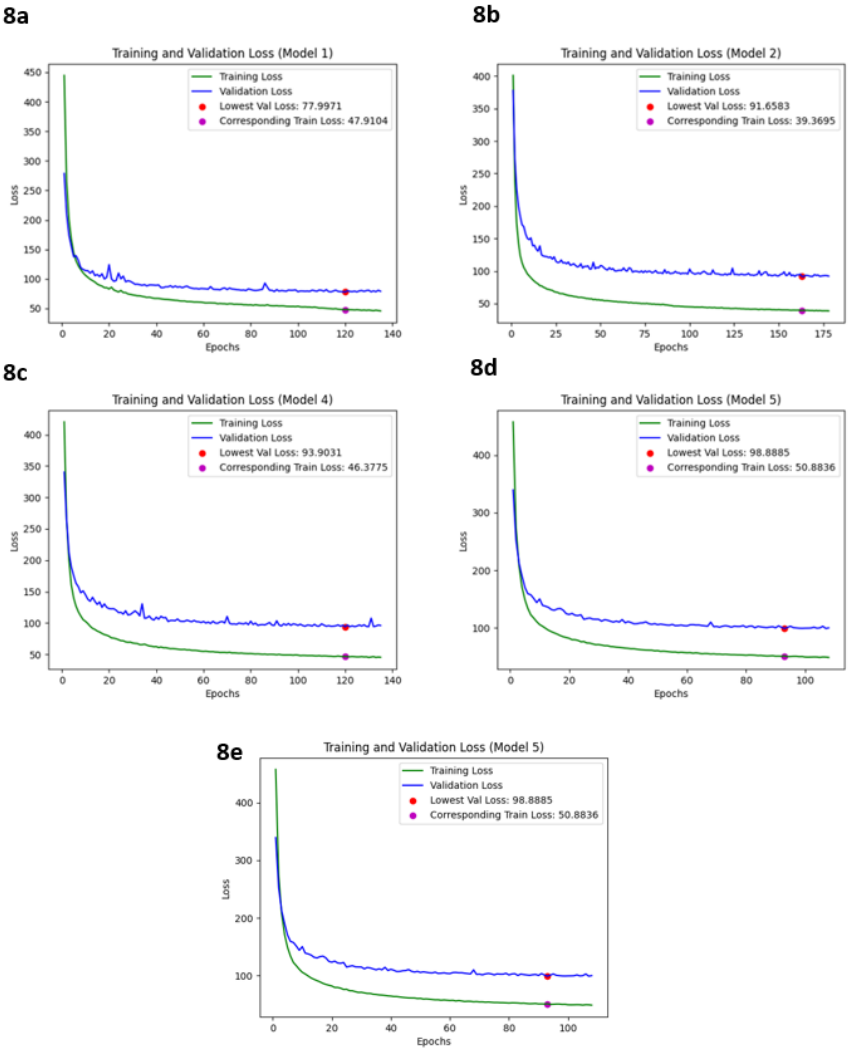

Figure 8. This shows 5 plots (a,b.c,d,e) representing the train and validation loss curves for an autoencoder model with the goal of reconstructing the same ECG the had been inputted in the model. The loss is based on mean absolute error and each plot , plots the recorded MAE for each epoch. The best val loss and the corresponding train loss of the model is stated in the legend as this is the train and val loss of the model used

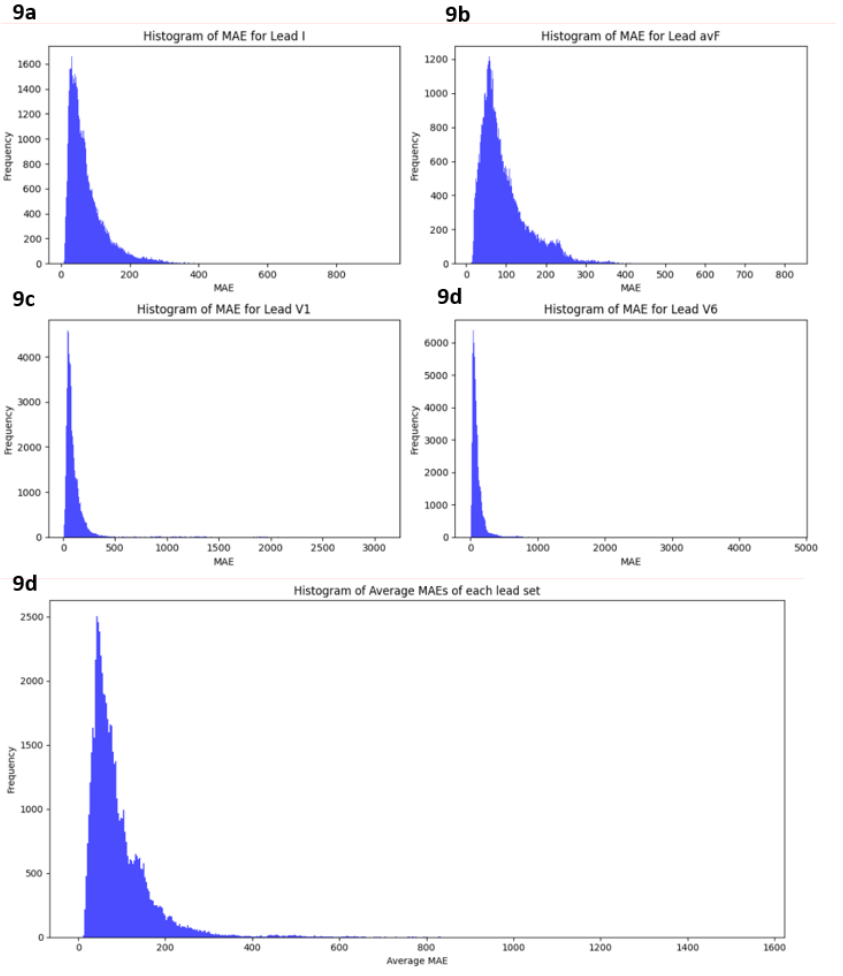

Figure 9. This figure shows the distribution of the MAE loss for each lead, a,b,c and d show the distribution of the MAE loss for each lead for all the test sets 1,2,3,4,5 so in total all the 62,100 samples for each lead. Figure 2e shows the distribution for the average MAE for each set of leads I,avF,V1 and V6 together for each of the 62,100 samples.

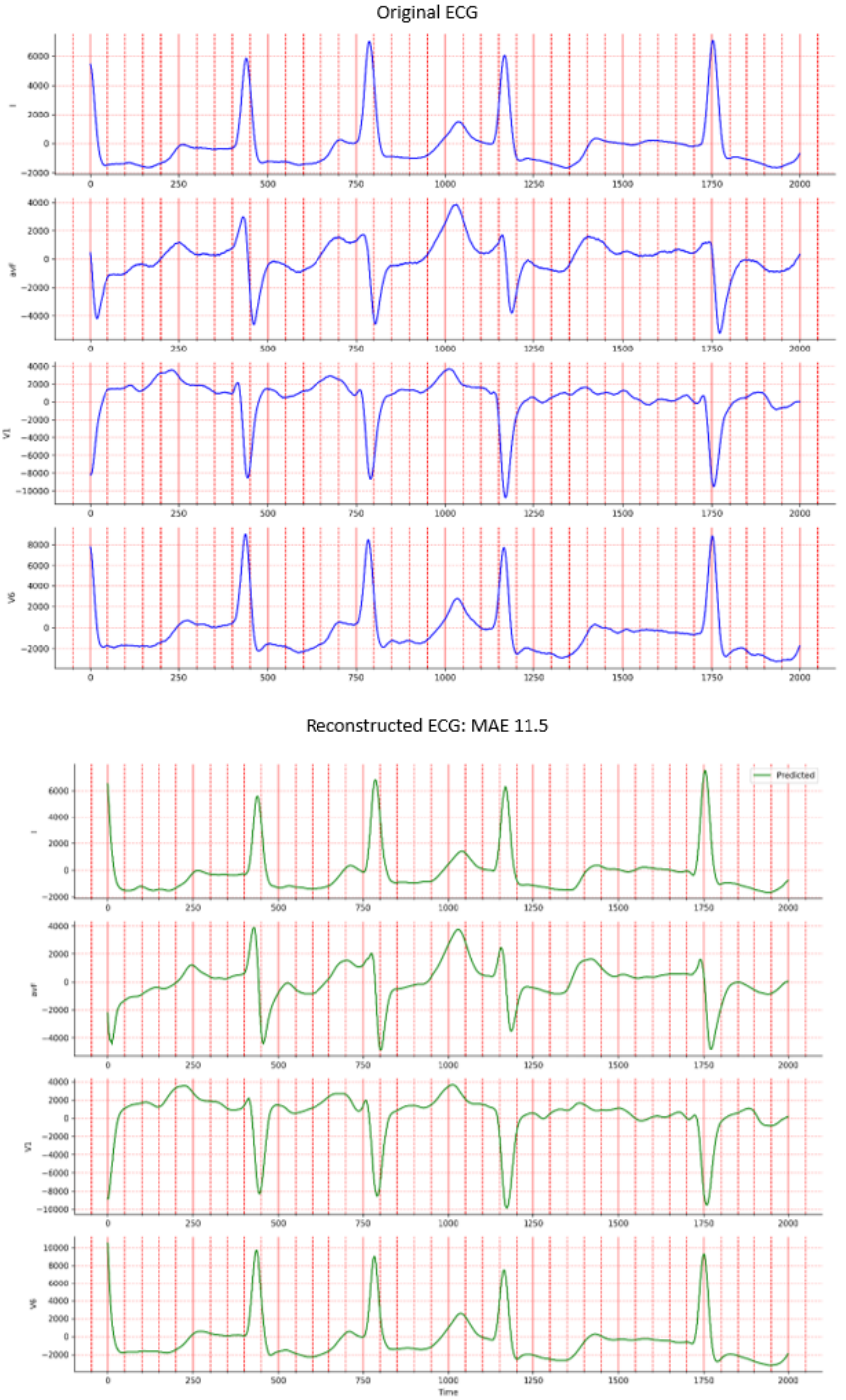

Figure 10. This plot shows what a good ecg reconstruction looks like . The plot are plotted in this order of lead I, avF,V1 and V6 .

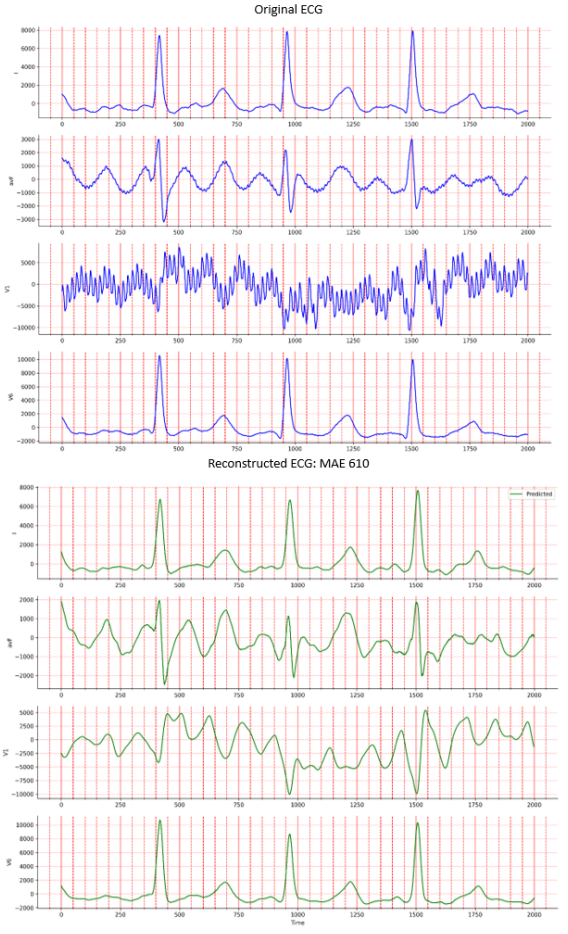

Figure 11. This shows an original and reconstructed ECG for a bad reconstruction with a high mean absolute error : 610.3.

4.2 Evaluating autoencoder intracardiac reconstruction

Given the adequate performance of the ECG reconstruction, we can now be confident to move onto the intracardiac reconstruction using the 4 lead ecgs to see if we can replicate the intracardiac lead CS 3-4 from out 4 lead ECGs , again these results will be the test results of all 62,100 samples as specified in the study pipeline(figure 4).

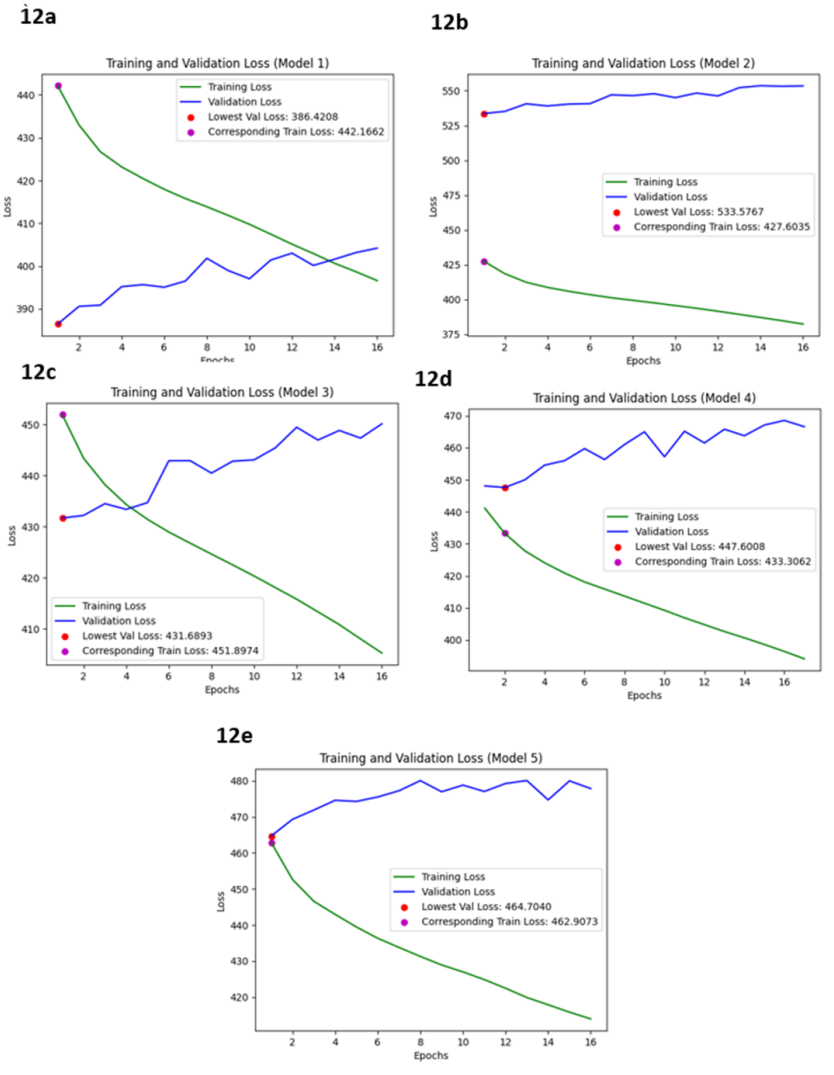

Figure 12. This shows 5 plots (a,b.c,d,e) representing the train and validation loss curves for an autoencoder model with the goal of reconstructing the intracardiac rhythm based on the ECG. The loss is based on mean absolute error and each plot , plots the recorded MAE for each epoch. The best val loss and the corresponding train loss of the model is stated in the legend as a best model was used based on the val loss therefore the model with the lowest val loss was being used even if the model lasted for more epoch. Each plot correspond to a model and each model tests for the corresponding fold e.g model 1 used to reconstruct for test 1 (fold1).

By looking at the train and val loss plots we can see that none of the models have performed adequately at all. In each model the more the model is fitted to the training set, represented by the decrease in training loss, the more the val loss disimproves(figure12). Looking at the distribution of the MAE most reconstructions fall between the 0 -1000 range with some outliers after the 1500 mark(figure13).For

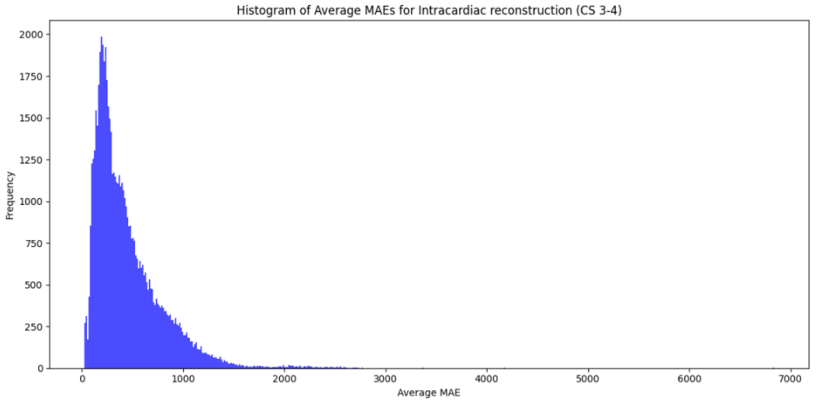

Figure 13. This figure shows the distribution of the MAE loss for each lead, here a,b,c and d show the distribution of the mae loss for each lead for all the test sets 1,2,3,4,5 so in total all the 62,100 samples . Figure 2e shows the distribution for the average MAE for each set of leads I,avF,V1 and V6 together for each of the 62,100 samples.

the good reconstruction we see how the placement of the QRS peak what is is being picked up well here however the details within that peak are inaccurate compared to the original signal also the erratic nature of the AF intracardiac reading before and after the QRS peak seem to not be captured at all. This a recurring theme among the well performing models where a smoothed out QRS is placed in the right place, but it does not capture any of the erratic nature of the intracardiac rhythm (figure 14). The predicted intracardiac that has a good MAE but performs badly visibly is interesting as the relatively small scale between -100 and 100 means that a MAE of 24.6 is actually a bad performance as opposed to a MAE of 24.6 at a larger scale (-400 – 400) which shows a better performing model (figure 15a). The final plot shows what a bad plot with a bad MAE (2598.0) actually looks like in which none of the QRS peaks are aligned properly and even the frequency of QRS peaks is not the same (figure 15b). It is also worth noting that the positioning of the QRS peaks is aligned with the peaks of the 4 leads influencing it rather than the intracardiac rhythm

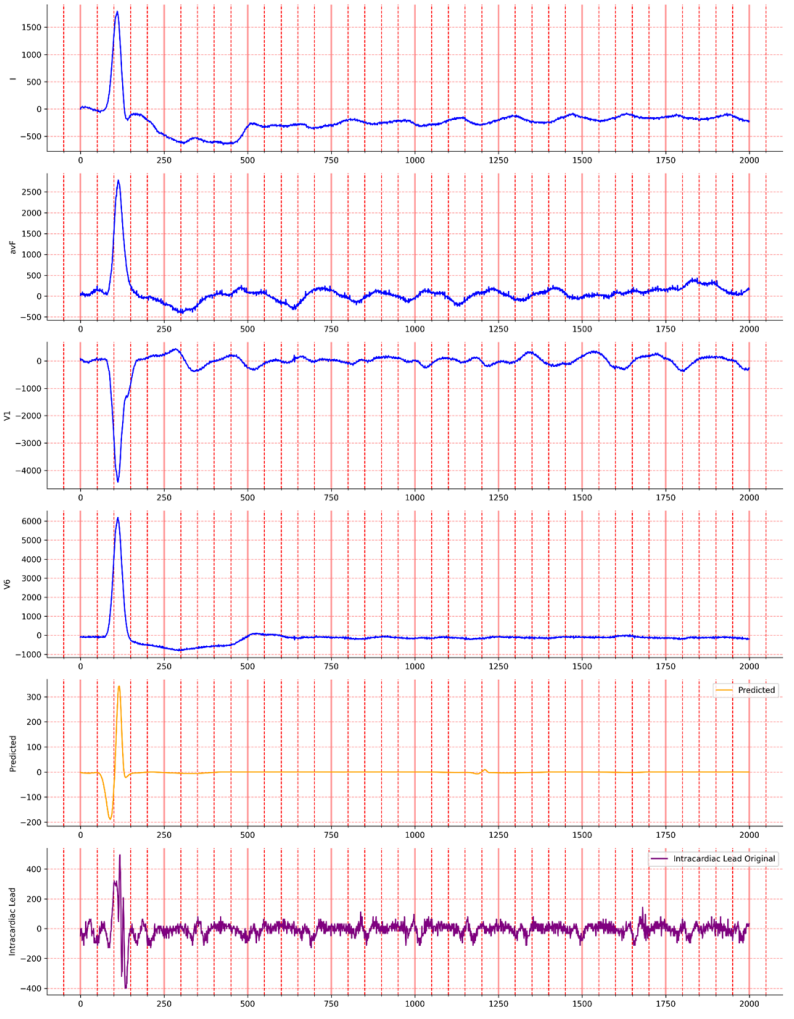

Figure 14. This is a well performing plot (MAE 24.64) . The plot shows the 4 ECG leads and the reconstructed intracardiac lead (predicted) and finally the original intracardiac rhythm plot is shows to compare with the predicted.

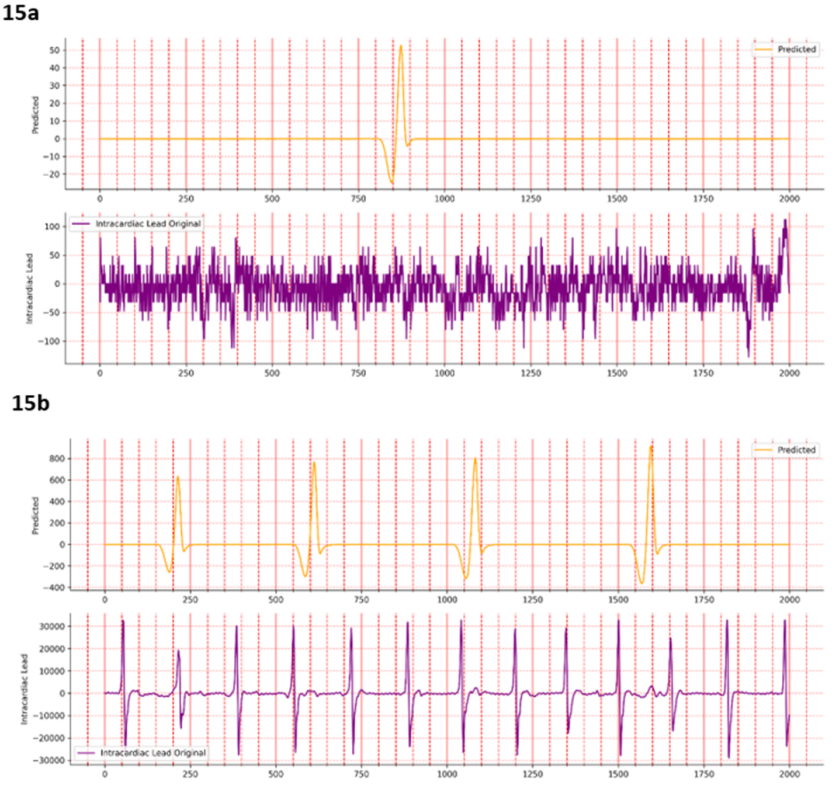

Figure 15. 15a shows a bad reconstruction visibly however the MAE is (24.6) and figure 15b just shows a poorly performing reconstruction visibly and based on MAE (2598.0)

4.3 CNN prediction performance of sample entropy

Since the intracardiac reconstruction from 4 ECG leads using an autoencoder has not met the conditions of any level of generalizability we decided to simplify the task at hand while still generating features related to the intracardiac morphology. We chose to predict the sample entropy as this seemed to adequately describe the complexity of the signal as the higher the sample entropy the more complex the signal. For this task a regression CNN would be required to predict the sample entropy of an intracardiac rhythm based on 4 ECG leads. To evaluate the performance of the CNN we will now use the train and validation loss curves and correlation plots for each test set.

Table 1. This table shows the performance of two activation functions used in the final layer of a model for predicting sample entropy. The model's evaluation is based on the correlation between predicted sample entropy (loaded predictions) and actual sample entropy values.

	Model 1	Model 2	Model 3	Model 4	Model 5	Total Performance
Sigmoid	0.26	0.24	0.39	0.26	0.23	1.38
Soft plus	0.21	0.15	0.29	0.23	0.21	1.09

The models seem to have an MAE that lies between 0.11 and 0.13, while most of the models do see a decrease in the validation loss as the model is fitted more to the train data Model 1 seems to not see any decrease and rather the val loss increases(figure 16). Model 1 also has many predicted values that hit a ceiling of 1, this is because the sigmoid function only outputs values from 0 to 1 (figure 17a). While a

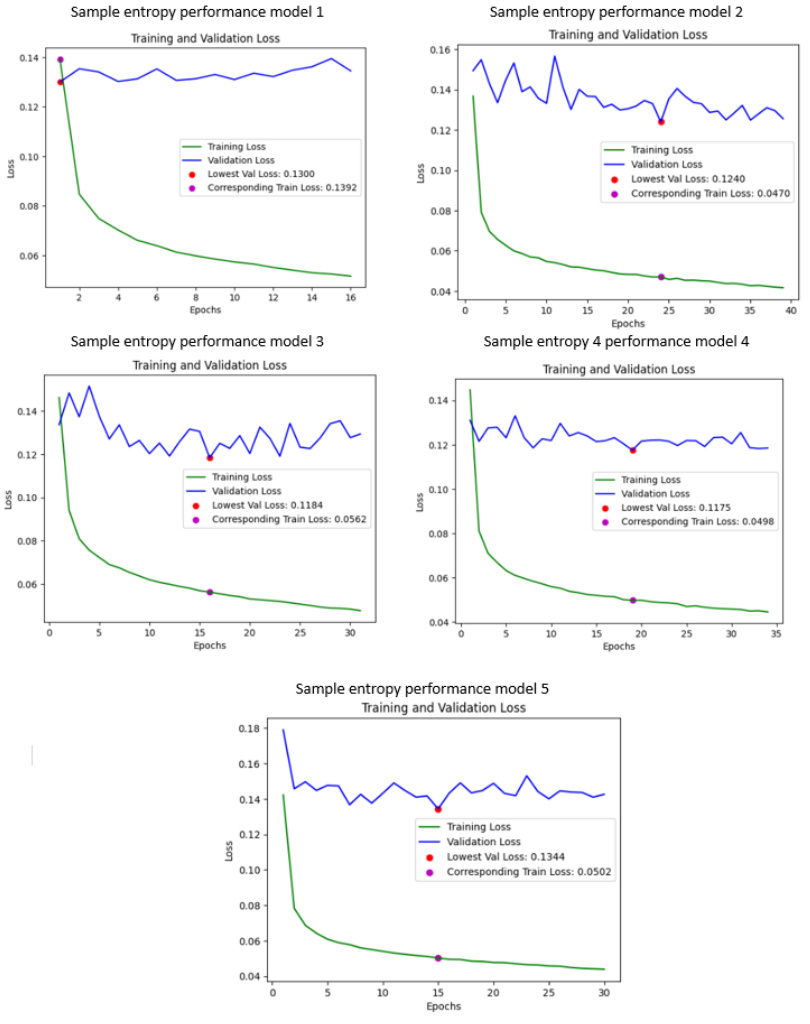

Figure 16. This shows 5 plots (a,b.c,d,e) representing the train and validation loss curves for an autoencoder model with the goal of predicting cross entropy of the intracardiac rhythm signal based on the ECG signal. The loss is based on mean absolute error and each plot , plots the recorded MAE for each epoch. The best val loss and the corresponding train loss of the model is stated in the legend as a best model was used based on the val loss therefore the model with the lowest val loss was being used even if the model lasted for more epoch. Each plot corresponds to a model and each model tests for the corresponding fold e.g model 1 used to reconstruct for test 1 (fold1). More details on the train val test split can be found in the methods flowdiagram(figure1)

soft plus function in theory should be better suited to this task as it can be trained and outputs and value that are positive, we opted to focusing on the sigmoid activation function as it performs better than the soft plus activation. This is seen in the total performance improvement of 0.29 over the total performance of the soft plus activation function(Table 1). By looking at the plots the best performance for the model occurs between 0 and 0.4 sample entropy values as seen by the larger accumulation of values around the optimal line compared to with the values go over 0.4 in which much less predicted values intersect with the optimal line (figure 10). Overall, the performance is quite poor for predicting sample entropy using ECG as the total performance is 1.38 out of the possible 5 and none of the correlations are over 0.4 which may result in a weak set of features to be extracted that are not indicative of effective features that can truly predict sample entropy from a 4 lead ecg reading. The clustering analysis will inform us of the effectiveness

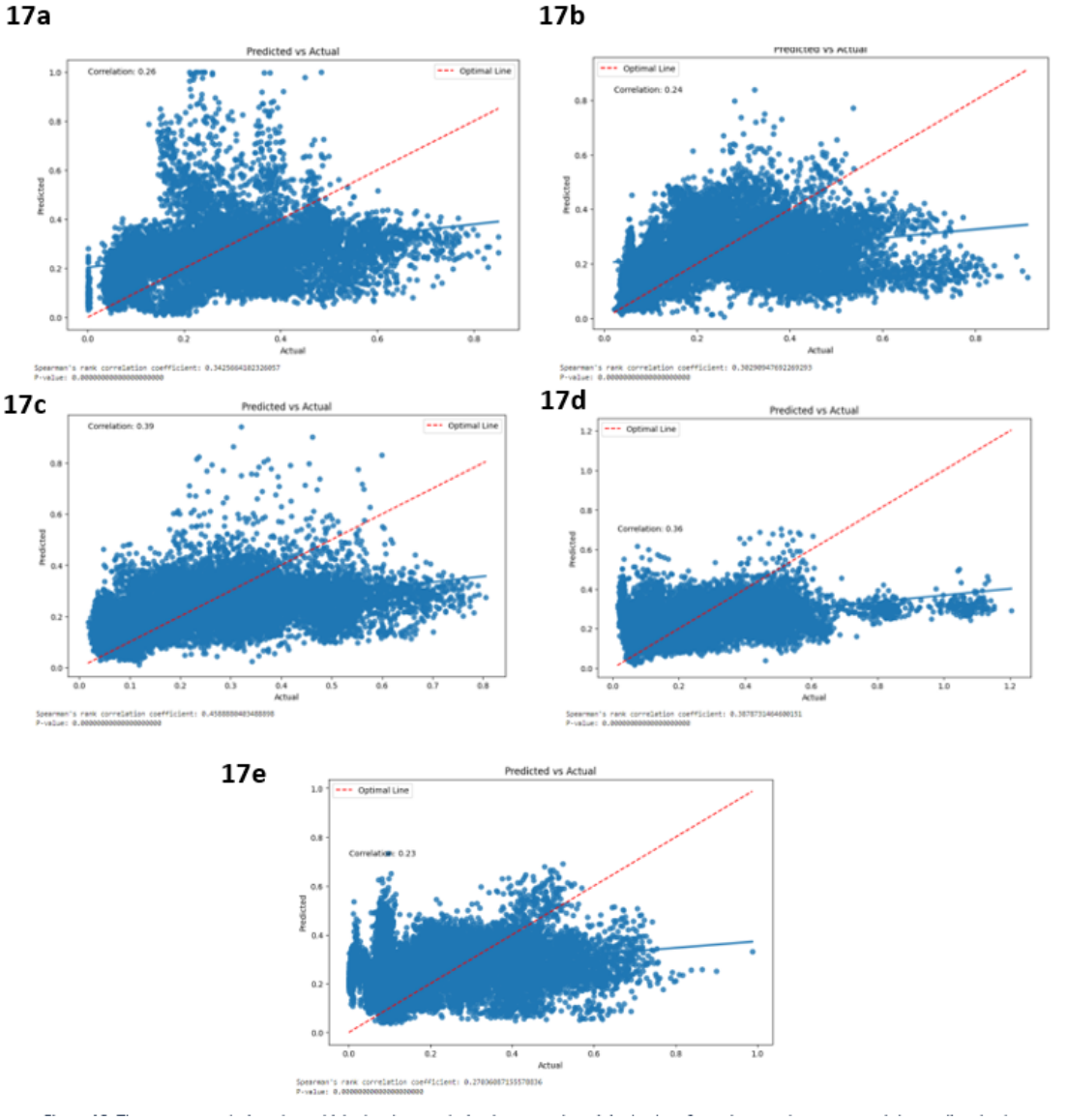

Figure 17. These are correlation plots that plot the correlation between the original values from the sample entropy and the predicted values for sample entropy. Plots a,b,c,d, and e all correspond to a model: 17a(model 1),17b(Model 2),17c (model3), 17d(model 4) , 17 e (model 5) and all these models used SIGMOID as the activation function in the final layer.. The ideal correlation would be 1 in which the loaded and predicted values are exactly the sample this is represented by the optimal line (red dashed line) The correlation for each model in order is 0.26,0.24.0.39.0.26,0.23. In Spearman rank p value is 0 for each which indicated all model perform better than a random guess of predicted sample entropies.

of these features in identifying unique characteristics across clusters.

4.4 Cluster analysis

4.4.1 Numerical variable analysis

Table 2. This table shows the mean and standard deviation from some of the numeric variables that where clustered on. The red boxes represent a Bonferroni corrected significance of 0.00714 so anything under this number is still considered significant. Some labels are not clear so I will detail them now eGFR_mlmin_173m2 stands for (eGFR(mL/min/1.73^2) which stands for the estimated glomerular filtration rate. LA here stands for the left atrium .LA dimension is the size of the atrium in mm .LA_area_2ch_2msq and LA_area_2ch_2msq are related to the atrial volume. LVEF_ is LVEF (%) which is Left ventricular ejection fraction percentage. CHADsVASc is a score that can suggest treatment based on a score up to 9. CAAPAF is a score for treatment as well. APPLE score is another score that is used to predict how likely a recurrence will help especially for patient with repeat ablation.

			Cluster				
Characteristic	1,N = 42,438	2 N = 37	3. N = 258	4. N = 6,203	5. N = 14	6 N = 107	7. N = 473
Age_Years	63 (9)	62 (10)	65 (9)	63 (9)	71 (12)	65 (8)	65 (11)
BMI	29.5 (6.3)	28.6 (3.0)	26.9 (5.2)	29.6 (5.7)	28.7 (2.2)	28.4(3.2)	30.0(5.2)
eGFR (mL/min/1.73m2)	70 (14)	76(13)	77 (9)	69 (14)	74 (5)	81 (11)	68(14)
LA Dimension (mm)	38.8 (6.1)	40.8(5.2)	35.3 (4.9)	38.9(6.2)	41.0(4.4)	38.6 (4.6)	38.0 (6.0)
LA Area 2Ch (cmsq)	23.56 (2.75)	23.03 (1.78)	22.49 (1.20)	23.57 (2.49)	25.29 (2.82)	22.46 (1.31)	23.69 (1.87)
LA Area 4Ch (cmsq)	25.2 (4.0)	25.3 (3.5)	23.6 (1.8)	25.2 (3.8)	27.1 (6.0)	23.3(2.9)	25.1 (2.8)
LVEF (%)	51 (7)	54 (2)	52 (5)	50 (7)	58(6)	52 (4)	50 (8)
CHADsVASc	3.00 (1.38)	2.08(0.72)	3.08(0.82)	3.11 (1.47)	2.80(0.11)	2.80 (1.40)	3.59 (1.54)
CAAPAF	3.92(1.53)	4.27(2.71)	3.77(1.49)	3.94(1.62)	4.36(1.45)	3.92(1.72)	4.01 (1.73)
APPLE	2.04 (0.90)	2.03 (0.83)	1.71 (0.58)	2.14 (0.94)	1.93(0.41)	1.79(0.70)	2.37 (1.02)
Atrial Diameter(cm)	1.63 (0.97)	1.84 (0.93)	1.23 (0.45)	1.67 (0.98)	1.95(0.58)	1.37(0.75)	1.58 (0.84)
sample entropy	0.27 (0.18)	0.20(0.08)	0.21(0.10)	0.26(0.16)	0.65(0.18)	0.13(0.09)	0.29 (0.17)

As per the secondary objective, we now have a set of features we have clustered on, statistical analysis is required to determine if there are any patterns within the characteristics of the clusters within our data. We can see how sample entropy seems to be significant in deciding cluster assignments in which cluster 5 seems to have the largest sample entropy values on average(0.65) and cluster. Given these results, it is worth breaking down the clusters that differ in sample entropy and the patterns seen in their characteristics.

Cluster 5

For cluster 5 which has a higher sample entropy on average(0.65), it is interesting to see that when compared to the first cluster the only things that are significantly different are age (71 which is the highest mean age), LVEF %(58) and atrial diameter(in which it has the highest mean out of all the groups for all three of those factors indicating that these are some of the most important factors in resulting in a higher sample entropy. The increased level of all these risk factors is linked to a greater risk of all cardiovascular outcomes not just AF.

Cluster 6

In contrast, cluster 6 shows significantly higher levels of egfr compared to cluster one which indicates better kidney function, and low la volume as indicated by low 2ch and 4ch values and finally a smaller atrial diameter. All these factors are related to reduced cardiovascular risk suggesting that a lower sample entropy (0.13) may be indicative of a relatively healthier cohort compared to a higher sample entropy.

Cluster 3

This sample entropy pattern is also seen with cluster 3 which also has a significantly lower sample entropy (0.21) compared to cluster one and has the lowest BMI(26.9), la dimension(35.3) and atrial diameter(1.23) between all the clusters.

Cluster 7

When looking at cluster 7, which has a significantly higher mean sample entropy (0.29) , we also see a higher risk cohort that has the highest BMI score(30.0), scored the highest in every risk score in which its

APPLE score (recurrence) at 2.37 and CHadsvas(cardiovascular risk score) at 3.59 are significantly higher than in cluster 1. It also has the lowest eGFR (68).

Cluster 2

Cluster 2, also has variable values that suggest a cardiovascular risk like a high atrial diameter of 1.84 However compared to cluster 1 there are no significant differences that stand out other than a higher eGFR and lower CHADsVASc .

Overall analayis

Overall while scores like CHADsVASc and APPLE are good differentiators of the clusters, the clusters at more extreme ends of these risk factors like 5 and 6 require looking at more of the risk factors like la volume, atrial diameter, LVEF and egfr to get a better understanding of what really differentiates these clusters from each other. It is now useful to look at the categorical variables which include comorbidities, drug and ablation treatment options to see if we can identify different treatment methods for this group. These results also have provided some optimism in recognizing that even with a subpar predictive performance we are still able to establish and identify clusters with differing characteristics to one another.

4.4.2 Categorical variable analysis

Looking at the drug history of the participants in all clusters it is evident that in general a low amount of participants have any drug history beyond taking beta blockers and artefact of cohort being composed off people with cardiovascular risk factors and comorbidities as a common medicine to deal with arrhythmias is beta blockers which are above 80 percent for any cluster. Another commonality from the clusters is that most people went through RD ablation rather than cryoablation with most clusters other than cluster 3 having over 90 per cent of samples had RF ablation.

Cluster 5 as we have seen before was characterised by having the largest atrial dimensions like diameter, and la dimension. These may be linked to the fact that cluster 5 has a significantly higher proportion of samples with valvular disease (71%) and an enlarged left atrial (LA) diameter 71%. However, cluster 5 does have lower levels of other comorbidities like hypertension and congestive heart failure (14%) Finally for cluster 5 100 percent of the sample are pre catheter ablation giving us an idea of what a higher risk cohort may look like before catheter ablation.

In contrast, **Cluster 6** contained samples with relatively lower values for those risk factors linked to atrial dimension and we see within this cluster that these samples have the lowest number of samples with valvular disease at 0 percent which is significantly less than cluster 1. There are other factors that show a reduced risk compared to cluster 5 like la volume being significantly more normal than cluster 1 as 50% of the sample had a normal la volume. Cluster 6 also has the highest level of samples that are linked to taking amiodarone at 67 percent in addition to beta blockers at 98 percent. As far as drug history this where cluster 5 had drug history only linked to beta blocker cluster 6 showed a significant amount of amiodarone usage ad 67 percent another medicine linked to reducing disease.

Cluster 3 had the lowest la dimension, atrial diameter and this is backed up here by the sample in this cluster having a significant amount of sample with left atrial diameter over 43 mm , low level of impaired egfr and a low level of stroke however the samples seem to have a high level of hypertension at 70 percent. This cluster had a significant amount of sample that underwent cryoablation at 23%. while most clustered heavily favoured RF. Cluster 3 has no AF after ablation and while not significant it does have the lowest level of recurrence of arrhythmia out of all the clusters.

Cluster 7 also showed a significantly higher sample entropy mean, a low LVEF percentage at 50%. This is seen in the subjective lv function being severe which is significantly higher in this cluster, and is the highest among all the clusters at 29 per cent. In terms of drug history for cluster 7 the levels of digoxin use and amiodarone are also significantly higher than cluster 1 . Overall we see how the pattern of sample entropy being linked to risk continued throughout the look at the categorical variables aswell.

Table 3. This table represents the percentage split for each categorical variable. The values highlighted in yellow indicate a significant difference between the distribution of the categorical value in that cluster compared to cluster 1

Characteristic	Cluster						
	1	2	3	4	5	6	7
Sex MF							
Female	26%	0%	22%	23%	14%	25%	37%
Male	74%	100%	78%	77%	86%	75%	63%
Persistent or longstanding persistent 1 year							
Longstanding	72%	100%	63%	71%	71%	73%	62%
Persistent	28%	0%	37%	29%	29%	27%	38%
Hypertension (yes)	52%	59%	70%	50%	14%	49%	63%
Diabetes (yes)	9.8%	0%	10%	12%	0%	1.9%	13%
Coronary Disease (yes)	16%	30%	14%	16%	0%	12%	14%
Valvular Heart Disease (yes)	18%	0%	5.4%	18%	71%	0%	27%
Vascular disease (yes)	6.3%	0%	0%	8.6%	0%	25%	16%
StrokeThromboembolism(yes)	11%	0%	1.6%	13%	0%	0%	8.0%
OSA (yes)	4.3%	0%	2.7%	4.1%	0%	0%	9.3%
Heart Failure (yes)	31%	0%	27%	35%	14%	25%	33%
Beta Blocker Not including sotalol (yes)	83%	100%	88%	84%	86%	98%	86%
Calcium Channel Blocker YN verapamildiltiazem only	17%	0%	10%	17%	29%	0%	19%
Fleicanide (yes)	1.5%	0%	0%	2.1%	0%	0%	0%
Digoxin (yes)	17%	0%	3.9%	16%	14%	0%	23%
Amiodarone (yes)	30%	19%	11%	29%	0%	67%	52%
Sotalol (yes)	3.9%	0%	12%	3.2%	0%	0%	4.2%
Cryoablation or RF							
Cryoablation	9.4%	0%	23%	9.5%	0%	1.9%	11%
RF	91%	100%	77%	90%	100%	98%	89%
Was this patient in AF after completed ablation (Yes)	7.7%	0%	0%	7.1%	0%	0%	4.4%
Recurrence of atrial arrhythmia (yes)	92%	100%	90%	92%	100%	98%	100%
Congestive Heart Failure (yes)	31%	0%	27%	34%	14%	25%	33%
Diabetes (yes)	9.8%	0%	10%	12%	0%	1.9%	13%
imPaired eGFR <60 (yes)	18%	2.7%	1.9%	22%	0%	1.9%	36%
Left Atrial Diameter <43mm	27%	51%	1.9%	28%	71%	23%	27%
LVEF >50	16%	0%	5.0%	17%	0%	0%	18%
Coronary Artery Disease	16%	30%	14%	16%	0%	12%	14%
Subjective LV function (Mild)	9.3%	0%	10%	12%	14%	25%	17%
Subjective LV function (Moderate)	13%	0%	5.4%	18%	0%	0%	0.8%
Subjective LV function (Normal)	67%	100%	73%	58%	86%	75%	53%
Subjective LV function (Severe)	10%	0%	12%	12%	0%	0%	29%
procedure							
post	44%	30%	44%	48%	0%	36%	37%
pre	56%	70%	56%	52%	100%	64%	63%
LA Volume ml (normal)	11%	49%	10%	9.3%	14%	50%	3.0%
LA Volume ml (mildly dilated)	58%	49%	84%	61%	14%	49%	81%
LA Volume ml (moderately dilated)	18%	0%	3.9%	17%	71%	0%	11%
LA Volume ml (severely dilated)	12%	2.7%	1.9%	12%	0%	1.9%	5.1%

5. Discussion

While the autoencoder intracardiac rhythm reconstruction performance was deemed inefficient enough for feature extraction there was still some promise when looking at the use of CNN to predict sample entropy, a much simpler task. The CNN task also resulted in a performance that was considered less than ideal with a correlation score of less than 0.5, in spite of this we are able to establish some unique clusters based on changes in the mean sample entropy of each cluster which is what these features were directed towards. We have decided due to limitations in patient numbers to look at cluster 1 as a reference cluster and while not ideal a per-sample analysis is still effective in identifying clusters in this scenario. By using cluster 1 as a reference cluster we can identify how some clusters differ based on sample entropy and what changes this results in. We begin to recognise that a lower sample entropy is correlated with lower risk in some variables, for example, a smaller less enlarged atrial diameter. In addition, these larger entropy clusters seem to have a larger percentage of samples related to higher risk factors and AF comorbidities like diabetes, congestive heart failure and valvular disease.

5.1 Phenotype Analysis

The first result that is of interest is that there is promise in establishing new phenotypes based on sample entropy and we have linked this to certain risk factors, comorbidities and in some cases drug history. By looking at our clusters there is a difference in risk factors when it comes to clusters with samples with a significantly different mean sample entropy. Only clusters 2 and 4 do not show significant changes in sample entropy and therefore are not necessarily the focus of our phenotype analysis.

Researchers have previously established that sample entropy is a good way of characterising the complexity of ECG to the point where they are able to differential between patients with AF and without, so it is useful to be able to add to this concept by differentiation patients in more detail when looking solely at people with high-risk AF(persistent or longstanding) (Horie et al., 2018). In this case we are comapring at a higher risk compared to a lower risk as this cohort as a whole is at a higher risk of AF than the average population

5.1.1 What differences in sample entropy may indicate about risk

The first clear difference between the higher-risk high sample entropy and the lower risk low sample entropy clusters is the difference in atrial conformation. Overall, there is a higher mean atrial diameter and a greater percentage of samples suffer from an enlarged atrial diameter of over 43mm including a dilated left atrial volume. This is the first sign that patients that have samples within this higher entropy range are at higher risk as an enlarged left atrial diameter has been linked to a higher level of AF recurrence after ablation which is what we see (Zoni-Berisso et al., 2014). . While the recurrence of arrhythmia is high in all the groups, the clusters with high sample entropy like cluster 5 have arrhythmia recurrence of 100 per cent after catheter ablation. Severe dilation of LA volume has also been seen at a higher rate in the higher sample entropy clusters . Previous research again has linked dilated left atrial volume to a higher risk of cardiac hospitalization for people who underwent atrial fibrillation indicating that if we can link someone to this through predicting sample entropy we can identify people at higher risk of rehospitalization and recurrence(Wen et al., 2022). We can then prepare for this by scheduling more checkups so we can identify potential tachycardias before it gets to the point of rehospitalization.

From these clusters we can establish two new phenotypes with lower sample entropy while both are characterised by relatively better atrial conformation, and significantly reduced levels of comorbidity what separates them from each is their BMI which is much higher in cluster 6 and their proclivity to drugs used for treatment where cluster 6 has a higher amount of beta blocker and amiodarone used than cluster 3. Out of the two clusters, it can be argued that cluster 3 is less at risk as it has a smaller atrial diameter and smaller la dimensions. We can also establish two new phenotypes using the higher mean sample entropy clusters which have a greater risk of tachycardia when looking at the fact that significantly more samples

for both clusters 5 and 7 have an enlarged left atrial diameter and valvular disease. What separates them is that cluster 7 has a higher percentage of samples with comorbidities (hypertension and congestive heart failure) and more people with an impaired glomerular filtration rate(egfr).

5.1.2 Drug history analysis

Something that stands out when looking at the clusters overall is the lack of a diverse drug history. Most patients in the clusters are assigned beta blockers which does make sense as they are deemed effective in bringing sinus rhythm to normal levels (Kühlkamp et al., 2002). However, while beta blockers seem to be the standard for preventing cardiomyopathy during AF they are less effective and may have adverse effects on patients with a high ejection fraction. This is something that we see in these clusters in which the mean LVEF % for each cluster is greater than so using beta-blockers may not be as effective and may cause adverse effects to the patients(Meyer and Lustgarten, 2023).
Another drug of interest is amiodarone which is linked to reducing the number of sites that require ablation during pulmonary vein isolation catheter ablation (Miwa et al., 2014). In cluster 6 which has the lowest sample entropy, we identified that amiodarone use was highest in this cluster furthermore being linked to lower sample entropy however AF recurrence was still high at 98 percent while non of the patients were in AF after the surgery. It is of note that there is another cluster with 0 percent AF after catheter ablation that has a low amount of amiodarone use so it may not be a significant factor causing a successful catheter ablation Also it clearly does not deal with the high AF recurrence. Other drugs like flecainide were used at a very low amount but there is no evidence by looking at the clusters or through previous research that this can reduce adverse events like tachycardia of hospitalization after AF (Hayashi et al., 2014).
Given the poor performance of drugs to prevent recurrence or to treat AF initially it is clear that there must be research to find new novel treatments to treat AF .

5.2 Evaluating the practicality of constructing new AF phenotypes

There are many hurdles to overcome when leveraging these new phenotypes in a healthcare setting. The first which shall be discussed later is that these methods are reliant on being able to generate a robust set of features that is able to predict sample entropy. In this case, we are not able to predict sample entropy sufficiently as we can see in the model performance. This raises the question if the clusters and results will change in a situation where the models perform more adequately therefore producing a set of features that can better predict sample entropy. This is not to invalidate the current results as we have extracted an outline related to risk and sample entropy, but more accurate performance can help us develop a more robust set of clusters that is generalizable to a larger population.
Another question is how this research can be used in the future. As of now the information available here can fully inform current healthcare treatment but the new phenotypes we have suggested can help inform researchers looking into the topic. The reason why I believe this data is not yet suitable for a healthcare setting is that most of the cohort still has a high level of recurrence in arrhythmia after catheter ablation so freedom of arrhythmia using the methods clustered here is not an option and we have not established a cluster that has a level of recurrence smaller than the other. In its current state, there is promise for healthcare applications like using someone ECG to predict their average sample entropy and if they have a sufficient number of samples that have a higher sample entropy than the average of 0.2 we can determine that they may have a higher risk for comorbidities like strokes and congestive heart failure where we can prepare for that as best we can by antiarrhythmic drugs. This promise is mitigated by the fact that we do not know what currently works, as we see in the cluster and research I have mentioned drugs like amiodarone and beta blockers which have not produced promising results, especially in the recurrence of AF. There is no highly successful solution for people with persistent AF. This is due to the fact that there is a lot of heterogeneity within AF patients and successful treatment are highly tailored to the individual rather than a subset of AF and this is where this research shows the most promise(McLeod and Gersh,

2010).
The potential of a more tailored but generalizable approach can be helped by the development of new phenotypes, being able to explain AF at a higher resolution can help identify specific treatments for specific people rather than just separate between persistent AF and longstanding (Vitolo et al., 2021). For example, research papers have found that sotalol can return 50 percent of people with chronic AF to normal sinus rhythm so it is not used as often(Southworth et al., 1999. If we are able to establish more phenotypes within AF we can potentially identify which group specifically is benefitting from this simply by calculating the sample entropy or by predicting sample entropy using the ECG reading available to us to extract that patient's features and determining which cluster they may fit in(Southworth et al., 1999).

5.3 Appraisal of Models

Throughout this paper we attempt two main models and approaches one being the autoencoder and the other being the CNN approach in this section we will evaluate the performance of both models and attempt to explain the discrepancies in the research. First let us look at the autoencoder, from the initial ecg reconstruction analysis we are able to establish insight into what the autoencoder is actually doing and what features it is picking up on. Through this, we recognised that the autoencoder can sufficiently pick up and replicate the signal within an ECG in addition to dealing with some noise. This is consistent with the theory behind autoencoder as we are only extracting the most important features and then decoding them back into the original output this results in a noise reduction effect that can be useful in denoising ECGs however in this case it resulted in the model missing the small inflections between QRS peaks. (Lin, Liu and Liu, 2023). Looking at the reconstruction of the ECG intracardiac EGM reading our performance takes a sharp dip as we are unable to successfully replicate the intracardiac signal at a high resolution. However, for some samples especially samples with a singular qrs peak, we are able to replicate the correct area for the qrs peak, as far as replicating the exact conformation of the peak we generate a smoothed-out version of the peak(figure 14). These results are somewhat consistent with research from Banta et al who were able to reconstruct segmented ECG and EGM QRS peaks, in which they are able to reconstruct QRS peaks from ECGs to EGM s and vice versa (Banta et al., 2021). In contrast to this paper, they attempted this reconstruction on both filtered ECG and EMGs and only the QRS peaks nevertheless, feature extraction from the EGM to the ECG was conducted in which they were able to use these in a multiclassification approach to identify 5 different ECG types at a 0.97 accuracy. It is of note that our attempt involved attempting to reconstruct the whole EGM signal as opposed to just the QRS peaks but there are many useful insights we can extract from their methodology Another useful method is the use of a correlation to identify better-performing reconstructions this may allow us to identify and optimize towards truly better-performing reconstructions as Figure 15b shows that a low MAE may not be the best representative of a well-preforming model. The loss function used by Banta et al was the Pearson correlation coefficient which could be more effective in analysis how well the intracardiac morphologies are captured (Banta et al., 2021). • Another observation from our train loss model performance is that the model does improve in the train loss so it can be successful when reconstructing the EGM signal from the ECG but when we try to use these weights to reconstruct the EGM for another set of patients seen in the val loss the model completely fails and the validation loss actually increase as the model fits to the training data. This is also seen in the Benata et al paper as models trained on the same patients achieved a correlation of over 0.9 for each patient however there is a drop off when models trained on all other patients are tested on a patient that the model was not trained on in which they achieved an average correlation 0.765. In this paper, we had 259 and when it comes to reconstruction of the full EGM reading we still struggle to do so at a sufficient level suggesting that even more data may be required to achieve a sufficient generalizable reconstruction performance.

5.4 The looming data issue (limitations and suggested improvements for analysis)

The data collected in this situation was collected by members of the electrocardiomaths department at Imperial College London. This allowed us to have access to a greater variety of patients in which we were able to obtain 62,100 2-second samples from 285 patients at 1000hz. The availability of patients is much larger than what was previously available which was 14 patients available with the MIT-BIH arrhythmia database (Moody and Mark, 1992). As far as the ECG and EGM this database should have been sufficient for the task as other reconstruction ECG tasks and the previous EGJM reconstruction paper used fewer patients at 15 (Matyschik et al., 2020). This leads us to conclude that the discrepancies in the model performance are due to preprocessing choices and model methodology choices rather than having a small database. However, recreating the EGM signal in full is a novel task and there is currently no solid exact comparison to compare the task undertaken in this paper to any other research so it could be plausible that a more complex task like this may require a larger dataset. The larger issue with the data comes from the data analysis and the data available for each patient when it comes to drug history and atrial conformation. We can see that we are left with 196 patients for clustering so we lose 89 patients due to lack of data. In addition to this, there are patients that have missing values for some of these variables. Most cases have missingness of under 30 per cent and therefore were deemed under the common threshold for imputation which is 30 per cent. In some other cases, important variables like LVEF percentage, la volume and recurrence of arrhythmia had missingness over the threshold of >45 per cent. However, it has been suggested that missingness per column is not the best measure for imputation and FMI (fraction of missing information) which takes into account information retained by auxiliary variables when calculating missingness is a more effective measure to use before imputation(Madley-Dowd et al., 2019). For this paper, we have based the imputation performance on OOB error which was deemed sufficient but using FMI could be a useful tool moving forward. There where also struggles when analysing the clusters. We have done a per-sample analysis and not a per-patient analysis. The main reason for this is if we were to place patients in clusters that they are most frequent every patient would belong in the first cluster. We attempted several other methods to ascertain per-patient results like clustering only on the median sample entropy value for each patient or placing patients in clusters where they have the highest percentage of occurrence per cluster. However, these methods produced uninterpretable results as none of the differences between the clusters was deemed significant. This is why we settled upon a per-sample approach because we were able to differentiate what type of patient was more likely to have a higher sample entropy compared to a lower sample entropy and what patient characteristics were more influential in causing the different means across the samples. The reason why patients overlap between clusters is due to the fact that every patient has features from ECG that are similar to each other and most patients have samples with a sample entropy of 0.2. What differentiates some patients from others is that a smaller set of patients have a higher or lower sample entropy for some of their samples and here we are able to extrapolate what the most common features within those clusters are.

6. Conclusion

The CNN approach of predicting sample entropy was deemed a much more realistic goal than the reconstruction of an intracardiac lead using an autoencoder. The primary objective of model prediction was not completely fulfilled as the CNN approach also outputted results that were not ideal. Despite concerns of not having a robust set of features due to the poor prediction performance, we have been able to identify 4 potential phenotypes linked to AF. Clusters 3 and 6 were characterised by a low sample entropy and reduced risk factors like atrial diameter. Clusters 5 and 7 were characterised by a high sample entropy and increased risk factors like a severely dilated left atrium. While there is much work to be done until these clusters can be applicable in a healthcare setting such as improving the predictive performance and cluster analysis, there is promise in their applicability to current research. Overall, research undertaken throughout this study provides a good step in being able to identify clinically relevant clusters based on the intracardiac rhythm in a non-invasive manner.

7. References

1. 1D convolutional neural networks and applications: A survey - ScienceDirect (no date). Available at: https://www.sciencedirect.com/science/article/pii/S0888327020307846 (Accessed: 29 August 2023).
2. Alzubaidi, L. et al. (2021) 'Review of deep learning: concepts, CNN architectures, challenges, applications, future directions', Journal of Big Data, 8(1), p. 53. Available at: https://doi.org/10.1186/s40537-021-00444-8.
3. Ashley, E.A. and Niebauer, J. (2004a) 'Conquering the ECG', in Cardiology Explained. Remedica. Available at: https://www.ncbi.nlm.nih.gov/books/NBK2214/ (Accessed: 11 August 2023).
4. Ashley, E.A. and Niebauer, J. (2004b) 'Conquering the ECG', in Cardiology Explained. Remedica. Available at: https://www.ncbi.nlm.nih.gov/books/NBK2214/ (Accessed: 10 August 2023).
5. Atmaja, B.T. and Akagi, M. (2021) 'Evaluation of Error and Correlation-Based Loss Functions For Multitask Learning Dimensional Speech Emotion Recognition', Journal of Physics: Conference Series, 1896(1), p. 012004. Available at: https://doi.org/10.1088/1742-6596/1896/1/012004.
6. Atrial Fibrillation ECG Review (no date). Available at: https://www.healio.com/cardiology/learn-the-heart/ecg-review/ecg-topic-reviews-and-criteria/atrial-fibrillation-review (Accessed: 11 August 2023).
7. Bai, Y. et al. (2021) 'Understanding and Improving Early Stopping for Learning with Noisy Labels'. arXiv. Available at: https://doi.org/10.48550/arXiv.2106.15853.
8. Bank, D., Koenigstein, N. and Giryes, R. (2021) 'Autoencoders'. arXiv. Available at: https://doi.org/10.48550/arX
9. Banta, A. et al. (2021) 'A Novel Convolutional Neural Network for Reconstructing Surface Electrocardiograms from Intracardiac Electrograms and Vice Versa', Artificial intelligence in medicine, 118, p. 102135. Available at: https://doi.org/10.1016/j.artmed.2021.102135.
10. Chieng, D. et al. (2022) 'Catheter ablation for persistent atrial fibrillation: A multicenter randomized trial of pulmonary vein isolation (PVI) versus PVI with posterior left atrial wall isolation (PWI) - The CAPLA study', American Heart Journal, 243, pp. 210–220. Available at: https://doi.org/10.1016/j.ahj.2021.09.015.
11. 'Clinical ECG Interpretation' (no date a) ECG and ECHO. Available at: https://ecgwaves.com/product/clinical-ecg-interpretation/ (Accessed: 30 August 2023).
12. 'Clinical ECG Interpretation' (no date b) ECG and ECHO. Available at: https://ecgwaves.com/course/the-ecg-book/ (Accessed: 30 August 2023).
13. 'Clinical ECG Interpretation' (no date c) ECG and ECHO. Available at: https://ecgwaves.com/course/the-ecg-book/ (Accessed: 30 August 2023).
14. Di Marco, L.Y. et al. (2013) 'Characteristics of atrial fibrillation cycle length predict restoration of sinus rhythm by catheter ablation', Heart Rhythm, 10(9), pp. 1303–1310. Available at: https://doi.org/10.1016/j.hrthm.20
15. 'ECG interpretation: Characteristics of the normal ECG (P-wave, QRS complex, ST segment, T-wave)' (2023) ECG and ECHO. Available at: https://ecgwaves.com/topic/ecg-normal-p-wave-qrs-complex-st-segment-t-wave-j-point/ (Accessed: 11 August 2023).
16. Escribano, P. et al. (2022) 'Preoperative Prediction of Catheter Ablation Outcome in Persistent Atrial Fibrillation Patients through Spectral Organization Analysis of the Surface Fibrillatory Waves', Journal of Personalized Medicine, 12(10), p. 1721. Available at: https://doi.org/10.3390/jpm12101721.
17. Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge - Giuseppe Lippi, Fabian Sanchis-Gomar, Gianfranco Cervellin, 2021 (no date). Available at: https://journals.sagepub.com/doi/ (Accessed: 30 August 2023).
18. Hayashi, M. et al. (2014) 'Three-month lower-dose flecainide after catheter ablation of atrial fibrillation', Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology: Journal of the Working Groups on Cardiac Pacing, Arrhythmias, and Cardiac Cellular Electrophysiology of the European Society of Cardiology, 16(8), pp. 1160–1167. Available at: https://doi.org/10.1093/europace/euu041.
19. Horie, T. et al. (2018) 'Sample Entropy in Electrocardiogram During Atrial Fibrillation', Yonago Acta Medica, 61(1), pp. 49–57.
20. Ikotun, A.M. et al. (2023) 'K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data', Information Sciences, 622, pp. 178–210. Available at:

https://doi.org/10.1016/j.ins.2022.11.139.
21. January, C.T. et al. (2019) '2019 AHA/ACC/HRS Focused Update of the 2014 AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society in Collaboration With the Society of Thoracic Surgeons', Circulation, 140(2), pp. e125–e151. Available at: https://doi.org/10.1161/CIR.0000000000000665.
22. Khan, F. et al. (2023) 'ECG classification using 1-D convolutional deep residual neural network', PLOS ONE, 18(4), p. e0284791. Available at: https://doi.org/10.1371/journal.pone.0284791.
23. Kiranyaz, S. et al. (2019) '1-D Convolutional Neural Networks for Signal Processing Applications', in ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8360–8364. Available at: https://doi.org/10.1109/ICASSP.2019.8682194.
24. Kiranyaz, S. et al. (2021) '1D convolutional neural networks and applications: A survey', Mechanical Systems and Signal Processing, 151, p. 107398. Available at: https://doi.org/10.1016/j.ymssp.2020.107398.
25. Klimek-Piotrowska, W. et al. (2016) 'Normal distal pulmonary vein anatomy', PeerJ, 4, p. e1579. Available at: https://doi.org/10.7717/peerj.1579.
26. Kostadinov, S. (2019) Understanding Backpropagation Algorithm, Medium. Available at: https://towardsdatascienc backpropagation-algorithm-7bb3aa2f95fd (Accessed: 14 August 2023).
27. Koulouris, S. and Cascella, M. (2023) 'Electrophysiologic Study Interpretation', in StatPearls. Treasure Island (FL): StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/books/NBK560784/ (Accessed: 30 August 2023).
28. Kühlkamp, V. et al. (2002) 'Use of beta-Blockers in Atrial Fibrillation', American Journal of Cardiovascular Drugs, 2(1), pp. 37–42. Available at: https://doi.org/10.2165/00129784-200202010-00005.
29. Kuznetsov, V.V. et al. (2021) 'Interpretable Feature Generation in ECG Using a Variational Autoencoder', Frontiers in Genetics, 12, p. 638191. Available at: https://doi.org/10.3389/fgene.2021.638191.
30. Kuznetsov, V.V., Moskalenko, V.A. and Zolotykh, N.Y. (2020) 'Electrocardiogram Generation and Feature Extraction Using a Variational Autoencoder'. arXiv. Available at: http://arxiv.org/abs/2002.00254 (Accessed: 28 August 2023).
31. Lane, T. (2018) 'Transposed Convolutions explained with… MS Excel!', Apache MXNet, 2 November. Available at: https://medium.com/apache-mxnet/transposed-convolutions-explained-with-ms-excel-52d13030c7e8 (Accessed: 29 August 2023).
32. Lin, H., Liu, R. and Liu, Z. (2023) 'ECG Signal Denoising Method Based on Disentangled Autoencoder', Electronics, 12(7), p. 1606. Available at: https://doi.org/10.3390/electronics12071606.
33. Lippi, G., Sanchis-Gomar, F. and Cervellin, G. (2021a) 'Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge', International Journal of Stroke, 16(2), pp. 217–221. Available at: https://doi.org/10.1177/1747493019897870.
34. Lippi, G., Sanchis-Gomar, F. and Cervellin, G. (2021b) 'Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge', International Journal of Stroke: Official Journal of the International Stroke Society, 16(2), pp. 217–221. Available at: https://doi.org/10.1177/1747493019897870.
35. Liu, H. et al. (2020) 'Using the VQ-VAE to improve the recognition of abnormalities in short-duration 12-lead electrocardiogram records', Computer Methods and Programs in Biomedicine, 196, p. 105639. Available at: https://doi.org/10.1016/j.cmpb.2020.105639.
36. Madley-Dowd, P. et al. (2019) 'The proportion of missing data should not be used to guide decisions on multiple imputation', Journal of Clinical Epidemiology, 110, pp. 63–73. Available at: https://doi.org/10.1016/j.jclinepi.2019.02.016.
37. Mahida, S. et al. (2015) 'Science Linking Pulmonary Veins and Atrial Fibrillation', Arrhythmia Electrophysiology Review, 4(1), pp. 40–43. Available at: https://doi.org/10.15420/aer.2015.4.1.40.
38. Matyschik, M. et al. (2020) 'Feasibility of ECG Reconstruction From Minimal Lead Sets Using Convolutional Neural Networks', in 2020 Computing in Cardiology. 2020 Computing in Cardiology, pp.

1–4. Available at: https://doi.org/10.22489/CinC.2020.164.
39. McLeod, C.J. and Gersh, B.J. (2010) 'A practical approach to the management of patients with atrial fibrillation', Heart Asia, 2(1), pp. 95–103. Available at: https://doi.org/10.1136/ha.2009.000596.
40. Meek, S. and Morris, F. (2002) 'Introduction. I—Leads, rate, rhythm, and cardiac axis', BMJ: British Medical Journal, 324(7334), pp. 415–418.
41. Meyer, M. and Lustgarten, D. (2023) 'Beta-blockers in atrial fibrillation—trying to make sense of unsettling results', EP Europace, 25(2), pp. 260–262. Available at: https://doi.org/10.1093/europace/euad010.
42. missForest citation info (no date). Available at: https://cran.r-project.org/web/packages/missForest/citation.htm (Accessed: 31 August 2023).
43. Miwa, Y. et al. (2014) 'Amiodarone reduces the amount of ablation during catheter ablation for persistent atrial fibrillation', Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology: Journal of the Working Groups on Cardiac Pacing, Arrhythmias, and Cardiac Cellular Electrophysiology of the European Society of Cardiology, 16(7), pp. 1007–1014. Available at: https://doi.org/10.1093/europace/eut399.
44. Moody, G.B. and Mark, R.G. (1992) 'MIT-BIH Arrhythmia Database'. physionet.org. Available at: https://doi.org/10.13026/C2F305. 45. Nattel, S., Bourne, G. and Talajic, M. (1997) 'Insights into Mechanisms of Antiarrhythmic Drug Action From Experimental Models of Atrial Fibrillation', Journal of Cardiovascular Electrophysiology, 8(4), pp. 469–480. Available at:https://doi.org/10.1111/j.1540-8167.1997.tb00813.x.
46. Njoku, A. et al. (2018) 'Left atrial volume predicts atrial fibrillation recurrence after radiofrequency ablation: a meta-analysis', EP Europace, 20(1), pp. 33–42. Available at: https://doi.org/10.1093/europace/eux013.
47. Oleszak, M. (2023) Autoencoders: From Vanilla to Variational, Medium. Available at: https://towardsdatascience from-vanilla-to-variational-6f5bb5537e4a (Accessed: 29 August 2023).
48. 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.
49. Poole, J.E. et al. (2020) 'Recurrence of Atrial Fibrillation after Catheter Ablation or Antiarrhythmic Drug Therapy in the CABANA Trial', Journal of the American College of Cardiology, 75(25), pp. 3105–3118. Available at: https://doi.org/10.1016/j.jacc.2020.04.065.
50. Proietti, R. et al. (2015) 'A Systematic Review on the Progression of Paroxysmal to Persistent Atrial Fibrillation: Shedding New Light on the Effects of Catheter Ablation', JACC: Clinical Electrophysiology, 1(3), pp. 105–115. Available at: https://doi.org/10.1016/j.jacep.2015.04.010.
51. Ramkumar, M. et al. (2022) 'Auto-encoder and bidirectional long short-term memory based automated arrhythmia classification for ECG signal', Biomedical Signal Processing and Control, 77, p. 103826. Available at: https://doi.org/10.1016/j.bspc.2022.103826.
52. Reddy, S.A. et al. (2021) 'Pulmonary vein isolation for atrial fibrillation: Does ablation technique influence outcome?', Indian Heart Journal, 73(6), pp. 718–724. Available at: https://doi.org/10.1016/j.ihj.2021.10.012.
53. Rocca, J. (2021) Understanding Variational Autoencoders (VAEs), Medium. Available at: https://towardsdatascier variational-autoencoders-vaes-f70510919f73 (Accessed: 28 August 2023).
54. Rousseeuw, P.J. (1987) 'Silhouettes: A graphical aid to the interpretation and validation of cluster analysis', Journal of Computational and Applied Mathematics, 20, pp. 53–65. Available at: https://doi.org/10.1016/0377-0427(87)90125-7.
55. Saito, Y. et al. (2023) 'Phenotyping of atrial fibrillation with cluster analysis and external validation', Heart [Preprint]. Available at: https://doi.org/10.1136/heartjnl-2023-322447.
56. Singh, A. and Ogunfunmi, T. (2021) 'An Overview of Variational Autoencoders for Source Separation, Finance, and Bio-Signal Applications', Entropy, 24(1), p. 55. Available at: https://doi.org/10.3390/e24010055.
57. Southworth, M.R. et al. (1999) 'Comparison of sotalol versus quinidine for maintenance of normal sinus rhythm in patients with chronic atrial fibrillation', The American Journal of Cardiology, 83(12), pp. 1629–1632. Available at: https://doi.org/10.1016/s0002-9149(99)00168-x.
58. Stekhoven, D.J. (2022a) 'missForest: Nonparametric Missing Value Imputation using Random Forest'.

Available at: https://cran.r-project.org/web/packages/missForest/index.html (Accessed: 31 August 2023).
59. Stekhoven, D.J. (2022b) missForest: Nonparametric Missing Value Imputation using Random Forest.
60. Stekhoven, D.J. and Bühlmann, P. (2012a) 'MissForest—non-parametric missing value imputation for mixed-type data', Bioinformatics, 28(1), pp. 112–118. Available at: https://doi.org/10.1093/bioinformatics/btr597.
61. Stekhoven, D.J. and Bühlmann, P. (2012b) 'MissForest–non-parametric missing value imputation for mixed-type data', Bioinformatics (Oxford, England), 28(1), pp. 112–118. Available at: https://doi.org/10.1093/bioinfor
62. Ugarte, J.P., Tobón, C. and Orozco-Duque, A. (2019) 'Entropy Mapping Approach for Functional Reentry Detection in Atrial Fibrillation: An In-Silico Study', Entropy, 21(2), p. 194. Available at: https://doi.org/10.3390/e21020194.
63. Vitolo, M. et al. (2021a) 'Clinical Phenotype Classification of Atrial Fibrillation Patients Using Cluster Analysis and Associations with Trial-Adjudicated Outcomes', Biomedicines, 9(7), p. 843. Available at: https://doi.org/10.3390/biomedicines9070843. 64. Wasser, T.E. (2014) 'Increased Accuracy of Distribution Based Missing Value Imputation: An Alternative to Mean Imputation in Real World Environment Survey Research', Survey Practice, 7(3). Available at: https://doi.org/10.29115/SP-2014-0015.
65. Wen, S. et al. (2022a) 'Association of Postprocedural Left Atrial Volume and Reservoir Function with Outcomes in Patients with Atrial Fibrillation Undergoing Catheter Ablation', Journal of the American Society of Echocardiography, 35(8), pp. 818-828.e3. Available at: https://doi.org/10.1016/j.echo.2022.03.016.
66. Wen, S. et al. (2022b) 'Association of Postprocedural Left Atrial Volume and Reservoir Function with Outcomes in Patients with Atrial Fibrillation Undergoing Catheter Ablation', Journal of the American Society of Echocardiography, 35(8), pp. 818-828.e3. Available at: https://doi.org/10.1016/j.echo.2022.03.016.
67. Wen, S. et al. (2022c) 'Association of Postprocedural Left Atrial Volume and Reservoir Function with Outcomes in Patients with Atrial Fibrillation Undergoing Catheter Ablation', Journal of the American Society of Echocardiography: Official Publication of the American Society of Echocardiography, 35(8), pp. 818-828.e3. Available at: https://doi.org/10.1016/j.echo.2022.03.016.
68. What does atrial fibrillation look like on an ECG? (no date). Available at: https://theheartclinic.london/conditions/at fibrillation/answerpack/atrial-fibrillation/atrial-fibrillation-faq/what-does-atrial-fibrillation-look-like-on-an-ecg/ (Accessed: 11 August 2023).
69. Wilcoxon Signed Ranks Test - an overview | ScienceDirect Topics (2023). Available at: https://www.sciencedirect.c and-dentistry/wilcoxon-signed-ranks-test (Accessed: 31 August 2023).
70. Wilcoxon Test: Definition in Statistics, Types, and Calculation (no date) Investopedia. Available at: https://www.investopedia.com/terms/w/wilcoxon-test.asp (Accessed: 31 August 2023).
71. Wilson, D. and Martinez, T. (2001) 'The need for small learning rates on large problems', in, pp. 115–119 vol.1. Available at: https://doi.org/10.1109/IJCNN.2001.939002.
72. Wolf, P.A. et al. (1978) 'Epidemiologic assessment of chronic atrial fibrillation and risk of stroke: The framingham Study', Neurology, 28(10), pp. 973–973. Available at: https://doi.org/10.1212/WNL.28.10.973.
73. Xia, Y. (2020) 'Chapter Eleven - Correlation and association analyses in microbiome study integrating multiomics in health and disease', in J. Sun (ed.) Progress in Molecular Biology and Translational Science. Academic Press (The Microbiome in Health and Disease), pp. 309–491. Available at: https://doi.org/10.1016/bs.pmbts.2020.04.003.
74. Zoni-Berisso, M. et al. (2014) 'Epidemiology of atrial fibrillation: European perspective', Clinical Epidemiology, 6, pp. 213–220. Available at: https://doi.org/10.2147/CLEP.S47385.