An AI-based Predictor of Atrial Fibrillation Recurrence Following Pulmonary Vein Isolation

Jakob Botvidsson
2024

Master Thesis
in Biomedical Engineering

Supervisor: Frida Sandberg

LUND
UNIVERSITY

CathVision
Faculty of Engineering LTH
Department of Biomedical Engineering

Acknowledgements

I would like to thank Cathvision for giving me this opportunity, and for their hospitality. I would like to thank Rune Paamand, especially, for his invaluable support and insights throughout this project. Rune has been a big source of encouragement and inspiration, and our talks leave me with a fascination of what the future has to offer.

A big thanks to Frida Sandberg for guiding me in the thesis, but also for being open and caring.

Abstract

Atrial fibrillation (AF) is the most prevalent arrhythmia, increasing with each year. AF is characterized by periods of rapid and irregular beating of the heart and increases the risk of stroke and heart failure. The disease often originates from ectopic beats generated by the pulmonary veins, activating the atria, creating short periods of AF. Without treatment the disease evolves, remodels the atria, and eventually develops into a permanent condition of AF. Treatment with pulmonary vein isolation (PVI) through ablation is commonly prescribed. The treatment is effective, but AF recurrence post ablation occurs regularly. Clinical success is dependent on atrial health which can be interpreted through the P-wave of an electrocardiogram (ECG). With the help of this information, a more personalized treatment can be implemented, leading to increased arrhythmia freedom. Moreover, outcome prediction on ECG data using machine learning (ML) has proven efficient, but often uses extensive, inefficient manual delineation of the P-waves.

Therefore this study aims to evaluate AF recurrence prediction capabilities for two models trained on automatically extracted P-wave characteristics, raw ECG and coronary sinus (CS) catheter electrograms, respectively. Furthermore, one model is evaluated by training and testing on synthesized ECG data. The proposed method, applied to datasets of 281 subjects in total, automatically delineates ECG signals of adequate quality, recorded before, during and after PVI ablation. Annotations from delineation are used in beat classification, as well as to extract P-wave features, separately. All data is balanced based on beats per patient and outcome label and subsets with training and testing data are created. A Random Forest (RF) classifier is trained on the P-wave features and a Convolutional Neural Network (CNN) is trained on segments consisting of raw ECG and CS catheter electrograms. In order to further evaluate the CNN, a separate, synthesized dataset, to be used in training and testing, is created by inserting augmented P-wave characteristics in the ECG signals of AF recurrent subjects. The Area under the ROC Curve (AUC)-scores of the RF classifier and CNNs are 0.502 and 0.463. Furthermore, the AUC-score for the CNN training on synthesized ECG data is 0.793. Conclusively, neither of the RF classifier or the CNN can predict AF recurrence using this method and dataset. However, the result for the CNN training on synthesized ECG data illustrates the potential of a CNN to extract information from P-wave morphology. Moreover, val-

idation on automatic delineation on this dataset is needed to investigate the potential of the RF classifier. Conclusively, more research is needed on ML models trained on automatically delineated data.

List of acronyms & abbreviations

AF, Atrial Fibrillation;

PVI, Pulmonary Vein Isolation;

SR, Sinus Rhythm;

ECG, Electrocardiogram;

AP, Action Potential;

CS, Coronary sinus;

EP, Electrophysiology;

PWD, P-wave duration;

LA, Left atrial;

ACM, Atrial cardiomyopathy;

AI, Atrificial Intelligence;

ML, Machine Learning;

CNN, Convolutional Neural Network;

Contents

Acknowledgements **1**

Abstract **2**

List of acronyms & abbreviations **4**

1 Introduction **7**

2 Background **9**
2.1 Anatomy and Physiology of the Heart 9
2.2 The Cardiac Conduction of the Heart 11
2.3 Electrocardiogram 13
2.3.1 Coronary Sinus Catheter Electrograms 14
2.3.2 Automatic Delineation 15
2.4 Atrial Fibrillation 17
2.4.1 P-wave Morphology 18
2.4.2 Atrial Fibrillation Recurrence Post Ablation 19
2.5 Machine Learning 20
2.5.1 Random Forest 20
2.5.2 Neural networks 20
2.5.3 Convolutional neural networks 22
2.5.4 k-Fold Cross-Validation 24

3 Method **25**
3.1 Datasets 27
3.1.1 Ethics 28
3.2 Pre-processing 29
3.2.1 Visual assessment 29
3.2.2 Beat detection & classification 31
3.2.3 P-wave feature extraction 32
3.3 Machine Learning 34
3.3.1 Clinical data 34
3.3.2 A Synthesized Dataset 34
3.3.3 Random forest Classifier 36
3.3.4 Convolutional Neural Network 36
3.4 Model Evaluation 38

3.4.1 Evaluation Metrics 38
3.4.2 k-Fold Cross-Validation 38

4 Results **39**
4.1 Pre-processing 39
4.1.1 Visual Assessment 39
4.1.2 Beat classification 39
4.1.3 Statistical Analysis of Features 41
4.2 Prediction of AF recurrence 42
4.3 CNN Model Evaluation 43

5 Discussion **44**
5.1 Methods 45
5.2 Data 46
5.3 Survey of the field 47
5.4 Future work 48
5.5 Conclusion 50

References **51**

Appendix **63**

1 Introduction

There is a rise in cases of atrial fibrillation (AF) in the world population. In Europe, currently, the lifetime AF prevalence for subjects at age 55 is estimated to 1 in 3 individuals [1]. Many resources are allocated to combat the disease [2][3]. Treatment of AF with pulmonary vein isolation (PVI) through ablation has long been practiced in healthcare [4][5], and it improves the lives of many patients [6]. The treatment has a success rate of 32.9%, over 10 years, for patients with paroxysmal AF [7]. Despite its long history, the topic of clinical success post PVI is less understood [8][9]. For this reason, much of current research is devoted to exploring patient cardiac characteristics and tailored ablation strategies [10][11].

Research shows that left atrial (LA) remodeling correlates to clinical outcome of PVI [12][9]. Quantifying the extent of LA remodeling and defining the presence and location of low-voltage substrates (LVS) are used in targeted ablation of substrates along with PVI with clinical success [13][14]. But reliable ways of determining the extent of LA remodeling are invasive, intra-atrial mapping [15], or uses administration of gadolinium for magnetic resonance imaging (MRI) [16]. The LA remodeling in the form of LVS leads to an intra-atrial conduction slowing and can be measured as an increased LA activation time in intracardiac mapping [17]. An increased LA activation time is strongly correlated with a prolonged P-wave duration (PWD) in electrocardiograms (ECG) [17][18][19][20]. Similarly, P-wave amplitude indicates the extent of atrial remodeling, and thus is a predictor of AF recurrence [21]. Previously mentioned studies manually delineate and analyse the P-wave, which takes time and needs to be executed or reviewed by an experienced operator. Publicly available automatic ECG delineators perform on-par with expert annotators and can therefore speed up the process [22], however such automatic delineation has not previously been utilized in the context of predicting AF recurrence following PVI.

Machine learning (ML) has also shown potential in identifying patients at risk for arrhythmia recurrence post PVI ablation [23] and convolutional neural networks (CNN) are able extract information from time-series 1D ECG signals to detect AF [24]. It is, however, unclear whether there is potential in using a 1D CNN to analyse ECG signals and predict AF recurrence following PVI.

For this study three datasets are available, containing pre-, during and post-treatment recordings for 281 subjects that underwent PVI ablation. The recordings contain demographic information, ECG, CS catheter electrogram and treatment outcome.

Thus, the purpose of this study is to

- Investigate if automatic delineated P-wave characteristics can be used to predict AF recurrence in the present datasets.
- Investigate if a CNN can be trained to predict AF recurrence from raw ECG and CS catheter electrogram in the present datasets.

The report is divided into the following parts:

- **Background:** In Sec. 2.1-2.5, the reader will get the relevant background information on the subjects of anatomy and physiology of the heart, the ECG, AF and ML.
- **Method:** The datasets used is presented in Sec. 3.1. The methods of pre-processing through exclusion and inclusion criteria in visual assessment, beat classification, and P-wave feature extraction in Sec. 3.2. The last section, Sec. 3.3, presents data preparation for ML and how a synthesized dataset is generated, followed by implementation of ML models. The evaluation of models is described in Sec. 3.4.
- **Results:** Statistical values of pre-processing, P-wave feature distribution and measured model performance metrics are illustrated with tables and graphs in Sec. 4.
- **Discussion & Conclusion:** Results are discussed with regard to study limitations in methodology, in Sec. 5.1, dataset, in Sec. 5.2, and compared to successful implementations in previous research. Finally, a summary of the research field is presented in Sec. 5.3, followed by proposed future work in Sec. 5.4.

2 Background

2.1 Anatomy and Physiology of the Heart

The heart is a muscle situated in the center of the thoracic cavity, medially between the lungs. This area is called the mediastinum. The heart is separated from adjecent structures by a membrane called the pericardium. On the superior surface of the heart, the superior and inferior venae cavae, the aorta and pulmonary trunk attach to the heart. In the heart there are four chambers: the left and the right atria, and the left and the right ventricles. The left and the right side of the heart can be seen as separate pumps, where blood is pumped from the atria to the ventricles. For the right side of the heart, de-oxygenated blood is pumped to the lungs. In the lungs, gas exchange occurs, oxygenating the blood. The left side of the heart receives this oxygenated blood and pumps it to all the tissues of the body. The left and the right side of the heart is separated by the septum, and both the atria and the ventricles have outflow valves. Furthermore, the ventricles and outgoing arteries are also separated by a valve. These valves ensures one-way blood flow, opening from atria to ventricles and from ventricles to artery as the blood pressure rises with local muscle contraction.

Starting at the right atrium, de-oxygenated blood flows in from the superior and the inferior vena cavae. The atrium contracts, pumping the blood through the tricuspid valve into the right ventricle. As pressure decreases with the amount of blood increasing in the ventricle, the tricuspid valve closes. The ventricle contracts, pumping blood through the pulmonary valve out into the pulmonary trunk, which branches into the left and the right pulmonary artery going to the separate lungs. As blood flows through the pulmonary artery, the artery branches multiple times, until blood reaches the pulmonary capillaries. Here a gas exchange occurs where carbon dioxide leaves the blood and oxygen enters. As oxygenated blood flows towards the heart, the capillaries join together and forms the pulmonary veins. Four pulmonary veins connects to the heart, filling the left atrium with blood. The left atrium contracts, pumping blood through the mitral valve, into the left ventricle. With the same mechanical feature, the mitral valve closes and the left ventricle contracts, pumping the blood through the aortic valve into the aorta. Blood flow through the aorta, which branches multiple times as it goes

into the systemic arteries, branching even more as it enters the systemic capillaries. Here, an exchange occurs between capillaries and cells of the body as oxygen and nutrients exit the capillaries, to be used in the cell metabolic process, and carbon dioxide and waste products enter the capillaries. The de-oxygenated blood flows back towards the heart and, with capillaries joining together, entering the systemic veins and finally, joining together again, entering the major systemic veins, superior and inferior venae cavae [25].

The thickest tissue layer surrounding the chambers of the heart is the myocardium. The myocardium is what contracts to pump the blood into circulation. For efficient contraction, the cardiac muscles are wrapped circularly and spirally around each chamber. Like skeletal muscle, the contraction is driven by sarcomeres with sliding thick and thin filaments. Although, cardiac muscle cells are unique in that each cardiac muscle cell has a centrally located nuclei, and cells are connected through intercalated disks, thus creating a branched cellular network. Cell activation is driven by an electrical potential across the cell membrane, caused by a difference in ion concentration inside and outside the cell membrane. Across the cell membrane there are ion pumps and channels, allowing flow of ions across the membrane. A certain potential gradient is the action potential (AP). This depolarization causes sodium channels to open, further depolarizing the cell and opening of calcium channels. The calcium channels have a long opening duration allowing the cell to stay depolarized even though sodium channels are closed. The AP starts spontaneously in specialized cardiac cells and spreads across the muscles cells through interconnected gap junctions that allow free flow of ions, allowing an almost simultaneous activation of cardiac muscle [26].

2.2 The Cardiac Conduction of the Heart

In the heart, there are pacemaker cells that initiate the cardiac activation that propagates across the heart. The sinoatrial(SA) and atrioventruclar(AV) nodal cells are activated by a gradual rise in potential that eventually causes the calcium channels to open, depolarizing the cells [26]. The cardiac rhythm of a healthy heart is initiated by the SA node, located in the right atria. Although the activation of the SA node is automatic, it is affected by the sympathetic and parasympathetic nervous system. The purpose of the cardiac condution system is for the AP to propagate across the heart and activate cardiac muscle cells, this in an order that allows for the ventricles to be filled with blood from the contraction of the atria.

From the activation of the SA node the signal spreads across the right atria and towards the left atria and the AV node. In the right atria there are pathways for faster conduction to the AV node and to the left atria, called internodal tracts. The tracts are separated into three: the anterior tract, the middle tract and the posterior tract, all extending across the right atria and ending up at the AV node. The anterior tract also splits in two, with one pathway towards the AV node and the other into the left atria and the Bachmann's bundle. After atrial depolarization the signal reaches the AV node, located at the floor of the right atria, Figure 1. Signals reaching the AV node take either a slow or fast pathway. The slow pathway has slower conduction but a short refractory period and the fast pathway vice versa. At fast heart rates the slow pathway is used as the fast pathway is refractory, and in normal rhythm the fast pathway is used. From the AV node the signal is conducted to the bundle of His. The path to the AV node influence the time interval between atrial and His activation, with a longer interval for use of the slow pathway. Following the bundle of His the signal is split into the left and the right bundle branches, and onto the Purkinje fibers for each of the ventricles. Similarly to the tracts of the right atria, the Purkinje fibers allow for a particular rapid conduction. The rapid execution allows for a coordinated activation of the ventricular myocardium [27].

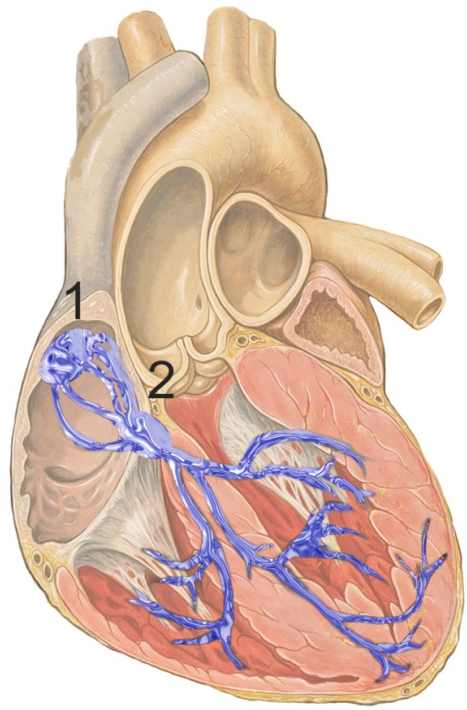


Figure 1: Conduction system of the heart. Normal AP starts from in the SA node (1) and continues through the atria and within internodal tracts (blue lines in right atria). AP reaches the AV node (2) and continues into the two separate purkinje fibers that branches across the ventricles [28].

2.3 Electrocardiogram

An electrocardiogram (ECG) measures the electrical activity of the heart. Cardiac muscle contractions are initiated by the AP that propagates throughout the heart and this synchronized depolarization of multiple cells is registered by the ECG. The AP propagation follows the cardiac conduction of the heart and as the cardiac cells of the atria and ventricles depolarize and repolarize, it creates a current flow in the heart. Two electrodes placed on the body is able to record the shifting electric field as the current flows through the heart. This is only possible since the body tissue is a conductive medium [29].

The electrical field measured between two electrodes is called a lead and the electrical signal is measured in units of mV. The standard in healthcare is a 12-lead ECG. There are six chest leads viewing the horizontal plane (V1-V6) and six limb leads viewing the vertical plane(I, II, III, aVR, aVL and aVF), see Figure 2. In the case of individual noise in a lead, using multiple leads, still allows cardiac information to be attained [30].

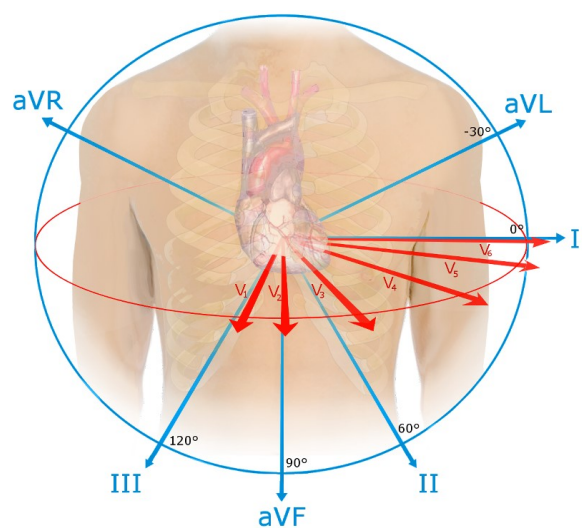


Figure 2: Spatial orientation of a 12-lead ECG [31].

For a normal cardiac cycle, the ECG contains a P-wave, QRS-complex and T-wave, see Figure 3. A positive amplitude in the ECG is measured when the current propagates towards the lead, a negative amplitude is measured

as the cells repolarize. The P-wave is a representation of the depolarization of the atria, the QRS complex is a representation of the depolarization of the ventricles, and the T-wave is a representation of the repolarization of the ventricles [30].

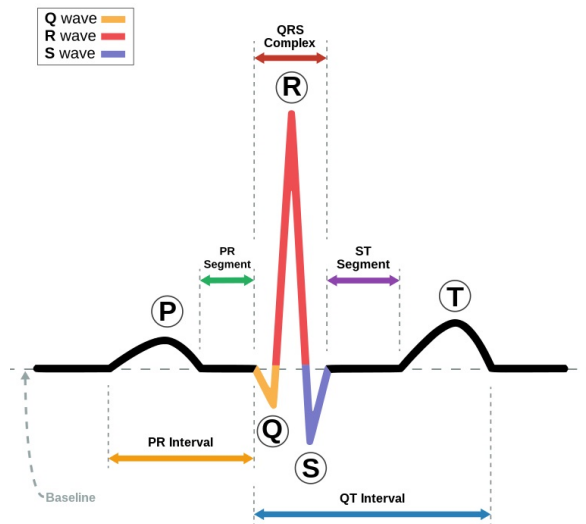


Figure 3: An individual heart beat, viewed as an ECG recording. P-wave, QRS-complex and T-wave are annotated [32].

2.3.1 Coronary Sinus Catheter Electrograms

The CS catheter electrogram is an intracardiac electrogram measuring cardiac electrical activation from the CS. The CS vein origin is between the left atrium and the left ventricle. Placing a catheter here, records both the atrial and the ventricular contractions. The catheter electrodes are located along the catheter, resulting in different recorded electrograms. The most proximal electrode therefore captures larger and more sharp atrial electrograms, and the distal electrode captures ventricular activation. In Figure 4, the pattern of proximal-to-distal activation is shown, where the first activation is picked up by CSp (coronary sinus proximal). Placement of the CS catheter therefore influence the activation sequence of the electrogram [33].

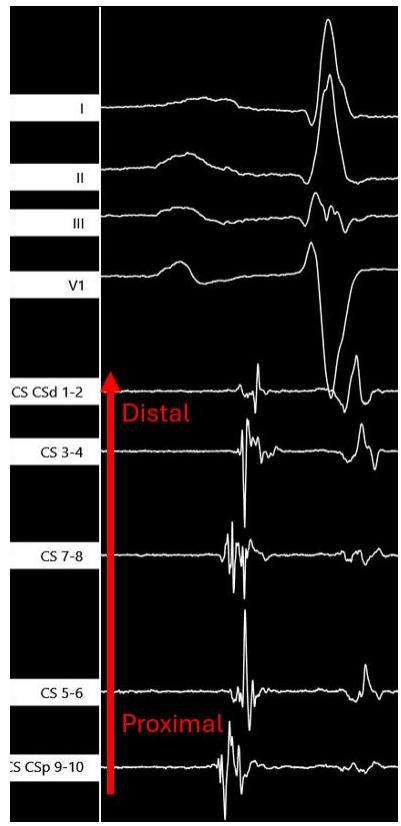


Figure 4: ECG of a single heart beat. CS catheter lead annotated proximal or distal placement of catheter lead electrodes in the CS vein.

2.3.2 Automatic Delineation

The ECG carries valuable cardiac information and the interval and amplitude measurements between carefully selected points in the ECG are used in diagnosis. An increase of smart devices in the last decades make it easier to accumulate large amounts of data [34]. Manual annotation is tedious and requires experienced operators. Automatically analysing ECG is therefore a resourceful way of improving diagnosis of cardiac disease. The QRS-complex is the most prominent feature of the ECG and, with the high amplitude, is also the easiest to detect. QRS-detection returns valuable information of R-peak to R-peak (RR) interval and therefore also the cardiac rhythm. The more difficult task is to determine peaks, onset and offset of the P-wave, QRS-complex and the T-wave, called delineation. QRS-detection is used in the delineation, where the location of the QRS is used to create adjacent windows of wavelet transforms. The windows are analysed to

detect P- and T-waves. The wavelet transform reveals the composition of frequencies, as a set of wavelets, in a signal [35]. P-wave delineation poses a challenge as the low amplitude is highly impacted by noise and that there is no clear rule to onset and offset of the wave [22]. Studies report, with use of various delineation approaches, delineation results on P-wave onset with errors(mean$\pm$SD(ms)) of -0.27 ± 12.2 [36], 2.00 ± 14.8 [22]. Delineation error can also occurs in manual measurements, as there is an observed variability, with intraobserver, $4.0 \pm 3.0\%$, and interobserver $5.0 \pm 3.5\%$ mean percent error [37].

2.4 Atrial Fibrillation

Atrial fibrillation (AF) is the most common cardiac arrhythmia. In Europe, currently, the lifetime AF prevalence for subjects at age 55 is estimated at 1 in 3 individuals. AF is recognized in an abnormal heart rhythm characterised by a rapid and irregular contraction of the atria. At an early stage of the disease, fibrillatory episodes can be short and rare, but as the disease progresses, the occurrence and duration of the episodes increase. The symptoms can be fatigue, chest pain and dizziness, but some patients are asymptomatic. AF is associated with an increased risk of stroke and heart failure [1]. There are plenty of established riskfactors for AF. Nonmodifiable risk factors are age, sex and genetics, while hypertension, diabetes mellitus, obesity and smoking are modifiable. These risk factors are related to inflammation, fibrosis and other forms of structural and electrical remodeling of the atria [38].

Traditional AF classification is based on presentation, duration and spontaneous termination of AF episodes [39]. AF can be *first diagnosed* or *recurrent*, if AF reappears after cardioversion/intervention, regardless of its duration or presence/severity. AF is classified as either *paroxysmal*, with AF episode termination spontaneously or with intervention within 7 days of onset, or *persistent*, with AF episodes sustained beyond 7 days. Persistant AF of more than 12 months in duration is instead defined as *long-standing persistent*. If no attemps are made to restore/maintain sinus rhythm and AF is accepted by both physician and patient, the AF is classified as *permanent*[39]. In the most recent AF guidelines, Joglar et al [40] argues that the traditional classification *"tended to emphasize AF once it was diagnosed and focused mainly on therapeutic interventions"* and proposes a new classification based on stages. This enables AF to be regarded as a progressive disease with different strategies at the different stages, *at risk for AF* looking at risk factors, *pre-AF* when there are structural or electrical findings which might predispose the disease, *AF* using the traditional classifications of paroxysmal, persistent and long-standing persistent and *permanent AF* defined as the traditional classification for permanent AF.

The mechanisms maintaining AF can be reentry or ectopic firing, called drivers. As AF is first presented in a patient it is often in paroxysmal form. Paroxysmal AF is usually initiated by ectopic beats in the cardiac muscle around the pulmonary veins. As the disease continues it evolves into Persis-

tant AF. This is due to the atrial remodeling caused by AF. Atrial remodeling can be structural or electrical and refers to any alteration in structure or function that promotes arrhythmia. remodeling causes development of reentry substrates, promoting Persistant AF. Treatment of AF can cause reverse remodeling, removing substrates, but if untreated these substrates can lead to Permanent AF. [41]

Atrial fibrillation can be treated with antiarrhythmic drugs to maintain sinus rhythm. This treatment involves side effects and difficulties keeping patients in long term sinus rhythm. A more successful treatment is catheter ablation, with the goal of isolating the cardiac tissue surrounding the pulmonary veins. The ablation scars the tissue, disabling conduction of ectopic activations into the atria. The procedure is often guided by visualizing the left atria anatomy and electrical activity using 3-D mapping. Ablation performed on patients with paroxysmal AF and minimal structural remodeling has a success rate of 60-75% [42]. Standard pulmonary vein isolation (PVI) ablation is not the only approach to remove AF, by increasing the target area for ablation, pairs of pulmonary veins are isolated in a method called wide-area circumferential ablation (WACA) [43]. Another approach of ablation is stepwise ablation, targeting drivers of AF in sequence. Targets are pulmonary veins, left atrial roof, left atrium (incorporating all anatomic regions of the chamber), mitral isthmus and non-LA structures [44]. Unfortunately AF recurrence post ablation occurs frequently, resulting in reablation for 20% to 40% of patients [45]. Reablation is not always the primary treatment for these patients considering risks associated with ablation, cost and disease burden of the patient. To make the right decision it is important to tailor the treatment to the personal characteristics of the patient [46].

2.4.1 P-wave Morphology

The ECG is a noninvasive tool for diagnosing atrial health. As the atria activates, the P-wave of the ECG visualizes the atrial conduction. The interpretation of the ECG can be linked to atrial function, fibrosis and interatrial blockage. A decrease in conduction between the atria is the definition of interatrial blockage, which is visualized in the ECG as a P-wave duration(PWD) > 120ms [47]. The presence of AF in a patient, remodels the tissue of the atria, slowing conduction. This is also visualized in the ECG as a prolonged PWD, which correlates with incidence of AF [48]. Left atrial enlargement is proposed to impact the P-wave area, and is correlated to an

increased risk of AF [49]. Determining the P-wave amplitude through the calculation of P-wave area and duration in lead II, also showed correlation between a flat-top P-wave and AF [50].

2.4.2 Atrial Fibrillation Recurrence Post Ablation

Research points to P-wave morphology when predicting risk of AF recurrence post ablation. Multiple studies show results of increased PWD correlation with arrhythmia recurrence post ablation. A study investigating the association of left atrial scarring in patients with paroxysmal AF found a prolonged PWD in patients prone to AF recurrence [51]. In research on PWD in pre-ablation SR ECG, a cut-off measure of >120ms to >150ms was associated with AF recurrence [18]. This allows for risk and treatment evaluation for patients before ablation, avoiding unnecessary risk for the patient and resources can be allocated to patients in need. A more accurate PWD cutoff value for arrhythmia recurrence is proposed by Jadidi et al [17], by manually measuring PWD in an amplified P-wave recorded prior to ablation, a cut-off value of 150ms is determined. Results showed an increased arrhythmia freedom for patients with PWD $<$ 150ms (72.7% vs. 30%). A ML approach to ablation outcome was investigated by Tang et al [52], developing a deep neural network leveraging ECG, electrogram and clinical data for prediction and a separate CNN using ECG alone. The ECG data used was pre-processed for the CNN to train on 5-sec segments with 12 channels from the 12-lead ECG. Area under the ROC Curve (AUC)-scores for the two models were 0.859 and 0.767 respectively. Jadidi et al [23] also proposed a ML approach for outcome prediction through classification of left atrial cardiomyopathy (ACM) using a neural network. Here a 12-lead ECG was annotated by experts, allowing feature extraction of PWD and P-wave morphology. The model performed well in classification of left ACM, which in turn allowed an ablation outcome prediction of 46% vs. 23% for patients with vs. without ACM.

2.5 Machine Learning

A ML model is adapting and learning from the data that it is being presented. As the ML model explores the data, a hypothesis of a pattern in this data is formed. When presented with new, unseen data, the model can make assumptions based on the hypothesis. Supervised learning is when the presented data is labelled with an answer. As the model makes an incorrect assumption, misaligned with the label, the model will have to change its hypothesis [53]. The data provided to the model is split into subsets for training and for testing. When the model is finished tuning to the pattern of the training data, it is presented with the unseen test subset and the predictions made by the model is then used to measure model performance. The size of the dataset provided to the model has a big impact on performance. Each data point carries individual information of a pattern and as a model is fed more data points it therefore has a more robust tuning to the data. One of the simplest examples of ML is linear regression, where a line is fitted to the data points so that the distance between each data point and the line is as short as possible. As more data points are added the line is realigned. When there are multiple patterns in the data, or the patterns are non-linear, models implementing more complex methods are needed.

2.5.1 Random Forest

A Random Forest (RF) uses multiple decision trees in classification. A decision tree is made up of nodes which split the data points based on cutoff values, starting at the root node and ending with subsets of data points in the leaf nodes. A RF implements multiple decision trees in a method called bagging, where multiple random samples of the data are used to train multiple decision trees. An average prediction across all trees is then used for classification by the RF classifier [54]. The use of multiple trees decreases the risk of overfitting and makes the model insensitive to noise. A RF has also shown good potential in classifying ECG beats leveraging features extracted using wavelet transform [55].

2.5.2 Neural networks

The design of the artificial neural network (ANN) is influenced by the biological neural networks in the human brain. This is a versatile and fast machine learning method that utilizes interconnected nodes similar to those

in the brain. Nodes process signals through weighted connections between themselves, with a distinct organization and interconnection pattern. The information moves across the neural network, from the input to the output. This is one of the simplest ANNs, called a Feedforward ANN since information moves in one direction [56]. An example can be seen in Figure 5.

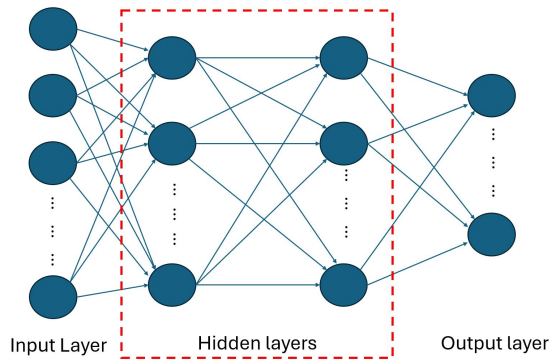


Figure 5: Feedforward ANN. Networks can contain different amounts of hidden layers.

The network learns by adjusting the weights between its nodes. During supervised training, these weights are optimized to minimize the error in predictions when compared to the provided labeled outputs. The mathematical function employed to reduce this prediction error is known as the loss function. The network's capability increases with the number of nodes, though extensive amounts of nodes in a single layer may compromise its ability to generalize to unseen data. By incorporating additional hidden layers, the network can be extended, creating what is known as deep learning. Deeper networks demonstrate enhanced generalization when trained on sufficiently large datasets, when trained on smaller datasets deeper networks can perform poorly and overfit to the data.

2.5.3 Convolutional neural networks

A CNN can be used for image recognition and object localization. By reducing images, they are made more manageable in processing while preserving essential features to be used in prediction. The CNN is similar to a feed-forward network but with three dimensions: width, height and depth. When the input is an image, the depth of the image are the features and could be the separate colors of red, green and blue. There are three types of layers in a CNN: convolutional, pooling, and Rectified Linear Unit (ReLU) layers. Each layer with different operations. The convolutional layer applies filters that propagate across the entire image. The dimensions of these filters and the number of filters used determine the dimensions of the next hidden layer, as shown in Figure 6. Each filter is designed to detect a specific feature in the image, often requiring a large number of filters to identify numerous features.

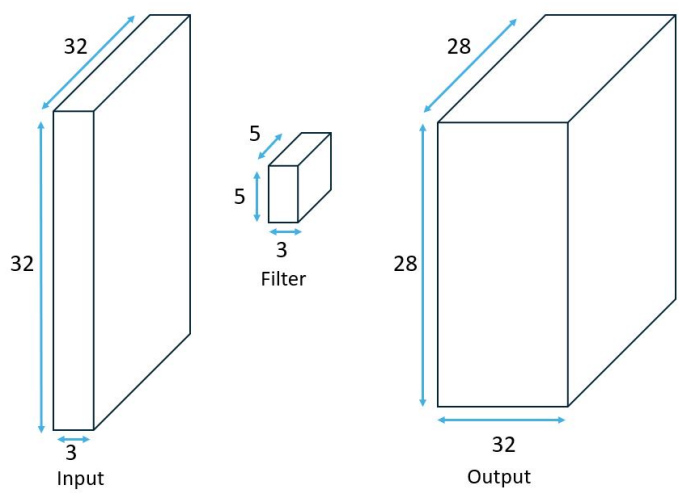


Figure 6: A convolutional operation between an input of size 32x32x3 and output of dimension 28x28x32 when applying a 32 filters of size 5x5x3. The filter and input must have the same depth. The output depth is determined by the number of filters applied.

The pooling operation reduces the size of the image while preserving the essential features in the image. Max pooling is a common pooling method that takes a small grid region from each layer of the image and outputs the maximum value from that grid, while maintaining the same depth. The ReLU

layer doesn't alter the image dimensions. It uses an activation function that sets negative values to zero. This increases processing speed, enabling the use of deeper models and generates more accurate networks. The ReLu layer also introduces nonlinearity to the neural network and is so commonly used that it is often not shown in network architecture diagrams [57].

In the present study a one-dimensional (1-D) CNN is used. 1-D CNN uses the same layers and operations, the difference is in the dimensions of the input, output and filters, that consist of the depth and only one spatial dimension, instead of the two in the 2-D CNN. 1-D CNN are applied to 1-D signals like ECG or accelerometer signals. The filters only move in the spatial dimension, on all feature channels. In the case of a 12-lead ECG signal of 1000 samples, the dimensions of the input would be 1000x12, that is 1000 samples of data for 12 channels. The filter would move in the time direction, across samples, for each lead of the 12-leads. The convolutional filters extract temporal information from the data which is well suited for extracting local features like frequency, amplitude and shape in signals. The CNN is also able to detect patterns invariant to the location in the signal [58]. 1-D CNN has proven efficient in detecting arrhythmic beats [59] as well as detecting AF [60].

2.5.4 k-Fold Cross-Validation

k-fold cross-validation is used to get more robust performance metrics when validating a ML model. Given a dataset, the model will train and test on a subset of that data. By resampling the subset for train and test, it's possible to generate unique subsets of testing data. k-fold cross-validation implements this by resampling the dataset 'k' times (see Figure 7), training and testing a model each time. The result is an average accuracy across all models.

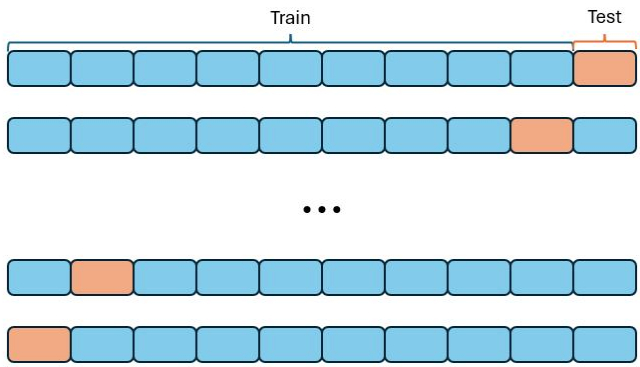


Figure 7: 10-fold cross-validation.

3 Method

Cathvision software RecorderApp is used for ECG visualization. Visual Studio Code is used as code editor with the Python programming language. Pandas and numpy are adopted for data handling, seaborn and matplotlib for visualization, and keras and scikit-learn are adopted for machine learning.

The proposed method for investigating prediction of AF recurrence follows the block diagram shown in Figure 8. Firstly, each patient's ECG recording is visually assessed to find 60s segments of adequate signal quality, as described in Sec. 3.2.1. The segments are automatically delineated to detect R-peaks and P-wave fiducial points as described in Secs 3.2.2 and 3.2.3. P-wave and R-peak annotations are saved and processed in separate methods. Exclusion and inclusion criteria are applied to R-peak annotations, generating a dataset of raw ECG and CS catheter electrogram data and labels, as described in Sec. 3.2.2. P-wave annotations are used in feature extraction, generating a dataset of features and labels as described in Sec. 3.2.3. Both datasets are balanced both in regards to data per patient but also in regards to labels, as described in 3.3.1. To further evaluate the capabilities of the CNN, a synthesized dataset is created using the raw ECG dataset, augmenting P-wave characteristics in the ECG for AF recurrent subjects, as described in Sec. 3.3.2. ML models are developed as described in Secs 3.3.3 and 3.3.4. The separate datasets are split into train and test subsets and the models are trained and evaluated, as described in Sec. 3.4.

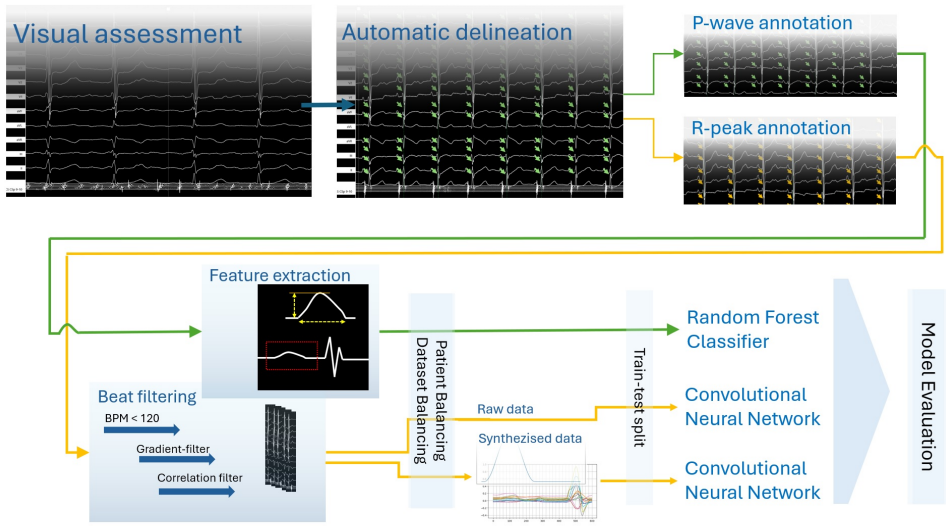


Figure 8: Block diagram of the proposed method for investigating AF recurrence predictors. In order, visual assessment (Sec. 3.2.1), automatic delineation, P-wave and R-peak annotation (Secs 3.2.2 and 3.2.3), beat filtering (Sec. 3.2.2) and feature extraction (Sec. 3.2.3). Generating synthesized dataset (3.3.2) and implementation of ML (Sec. 3.3), ending with model evaluation (Sec. 3.4). BPM: Beats per minute.

3.1 Datasets

Three datasets containing 281 subjects were used in this study. All subjects underwent PVI treatment for AF. Details on the characteristics of the datasets are shown in Table 1.

Dataset	Characteristic	
1	Age (mean, range)	61, (41-78)
	Male (n,%)	38, (95%)
	Time (in months) since first AF diagnosis (mean, range)	8, (2-46)
	Follow-up duration (months)(mean, range)	42, (8-65)
2	Age (mean, range)	63, (25-83)
	Male (n,%)	45, (78%)
	Time (in months) since first AF diagnosis (mean, range)	6, (2-30)
	Follow-up duration (months)(mean, range)	24, (12-54)
3	Age (mean, range)	72, (57-89)
	Male (n,%)	81, (62%)
	Time (in months) since first AF diagnosis (mean, range)	-
	Follow-up duration (months)(mean, range)	12, (12-12)

Table 1: Characteristics of the datasets included in the present study.

Dataset 1 contains 39 subjects treated at Lausanne University Hospital, Switzerland, using a stepwise ablation method. 12-lead ECG and CS catheter electrogram was acquired using CardioLab EP (electrophysiology) system (GE HealthCare) [61] at a sampling frequenccy of 1000 Hz. ECG was recorded thoughout the ablation treatment procedures; before, during and after catheter ablation with typically 1-2 hours of data available per patient. First follow-up occurred at different times but on average at a 42 months follow up (SD: 16,6 months). 8 subjects maintained sinus rhythm and 31 subjects suffered arrhytmia reccurence.

Dataset 2 contains 58 subjects treated at Lausanne University Hospital, Switzerland, using a wide area circumferential ablation method. 12-lead ECG and CS catheter electrogram was acquired using CardioLab EP system (GE HealthCare) [61] at a sampling frequency of 1000 Hz. ECG was recorded thoughout the ablation treatment procedures; before, during and

after catheter ablation with typically 1-2 hours of data available per patient. First follow-up occurred at different times but on average at a 24 months follow up (SD: 10.9 months). 17 subjects maintained sinus rhythm and 41 subjects suffered arrhythmia recurrence.

Dataset 3 contains 184 subjects obtained through the HRCRS (Heart Rhythm Clinical and Research Solutions) CRO, USA. 12-lead ECG and CS catheter electrogram was acquired using LabSystem Pro, Bard EP system [62] at a sampling frequenccy of 1000 Hz. ECG was recorded thoughout the ablation treatment procedures; before, during and after catheter ablation with typically 1-2 hours of data available per patient. First follow-up occurred after 12 months. 87 subjects maintained sinus rhythm and 42 subjects suffered arrhythmia recurrence, 41 patients did not return for follow up, 14 patients are missing labels.

3.1.1 Ethics

The data used in this study has been anonymized and shared through a contract/agreement with Lausanne University Hospital and HRCRS. To protect patients included in this study, pseudonyms are used. All patients included has consented to their data being used in research. Studies were conducted following ethics approval. The data used in the study was stored on a company hard-drive that only the writer had access to.

3.2 Pre-processing

3.2.1 Visual assessment

For the present study, if obtainable, a segment of 60s was extracted from each patient's ECG recordings. All patients with available recordings were included. The segments were selected as early as possible in the recording, where patients were in sinus rhythm and there was adequate signal quality. Recordings were probed for obvious signs of SR like stable rhythm, similar RR-intervals and indication of P-waves. Example of SR can be seen in Figure 9 and of AF in Figure 10. The CS leads are in focus to be able to look through each patient's hours of ECG recordings at a reasonable pace. CS lead's have a clear pattern in SR, with a spike at atrial activation and flat line in all other cases, Figure 11 shows examples of included and excluded recordings based on CS lead's. Examples of inadequate signal quality can be found in Appendix. In cases of recordings with adequate signal quality that are less than 60s, but more than 20s, recordings were included. Cases of doubt in abnormal recordings are solved by a second assessment by more competent colleges, working with ECG analysis within CathVision.

Figure 9: Example of adequate signal quality. Patient with clear SR. Areas marked for **A**: stable rhythm, **B**: indication of P-waves and **C**: a single atrial activation per beat.

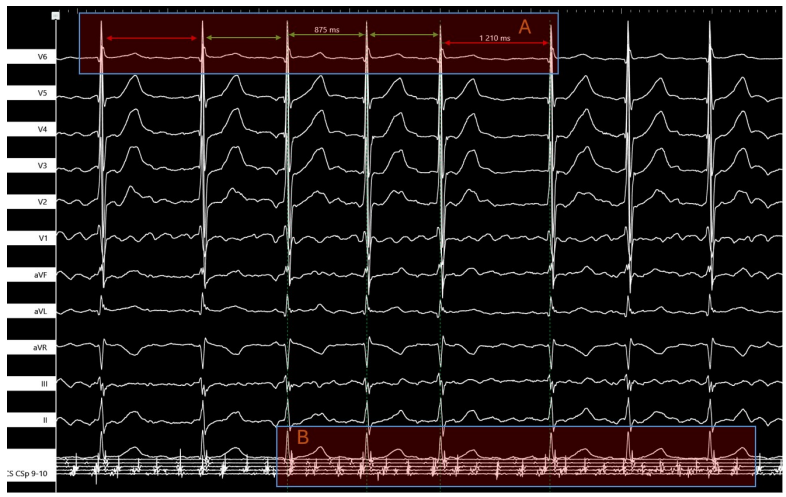


Figure 10: Patient with sample of AF, which can be seen in **A**: irregular rhythm and **B**: irregular activation of the atria.

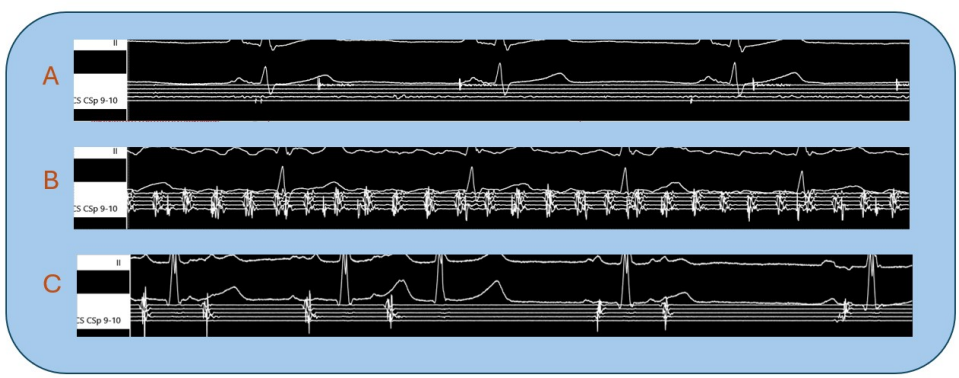
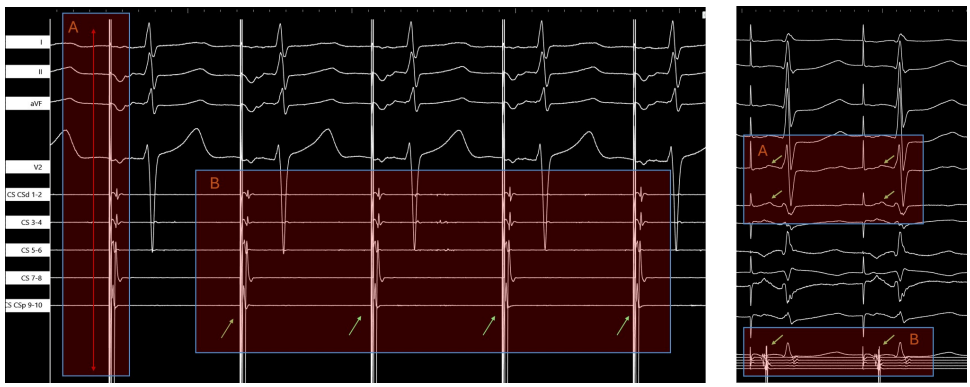


Figure 11: Catheter leads with **A**: no activation data, **B**: non-SR and **C**: Premature atrial complex.

In this study, segments with paced rhythm was included as well as segments with some premature atrial or ventricular contractions. These segments were used if recordings did not contain any other 60s segment with clear SR. Different cases of paced rhythm exist in recordings and will be visualized differently in the ECG. Examples of excluded and included pacing is shown in figure 12.



(a) Excluded recording. (b) Paced SR.

Figure 12: Examples of recordings with pacing. (**a**) Excluded recording, exclusion criteria is **A**: catheter lead pacing spike amplitude $> 10mV$ and **B**: single atrial activation per beat at the same location as pacing stimulus in catheter leads. (**b**) Included recording, **A**: indication of P-wave and **B**: pacing and atrial activation not interfering.

3.2.2 Beat detection & classification

Processing is done for every segment, separately. Firstly, the segments are automatically annotated with indices for R-peak locations in lead I, using the `signalDelineation(sig, fs)` function [22]. Each detected R-peak is regarded as a heartbeat. Each heartbeat is cropped into P-QRS intervals defined 250ms before and 50ms after each R-peak. The segments are then processed in order to:

1. Exclude P-QRS intervals with T-wave overlap from the previous P-QRS interval. The RR-intervals are calculated from R-peak annotations and P-QRS intervals with adjacent RR-intervals below 500ms (120BPM) are excluded.

2. Exclude P-QRS intervals with pacing, the P-QRS intervals with a gradient above 0.15 $\frac{dV}{dt}$ in any lead are excluded. An example of an excluded P-QRS interval is visualized in Figure 13. This process is referred to as gradient exclusion in the results.

3. Exclude irregular or noisy P-QRS intervals, the median P-QRS interval for all non-excluded P-QRS intervals is computed using `np.median()` which is then used as a P-QRS interval template. For each P-QRS interval the correlation to the P-QRS interval template is computed and

P-QRS intervals with a correlation below 0.9 are excluded. This process is referred to as correlation exclusion in the results.

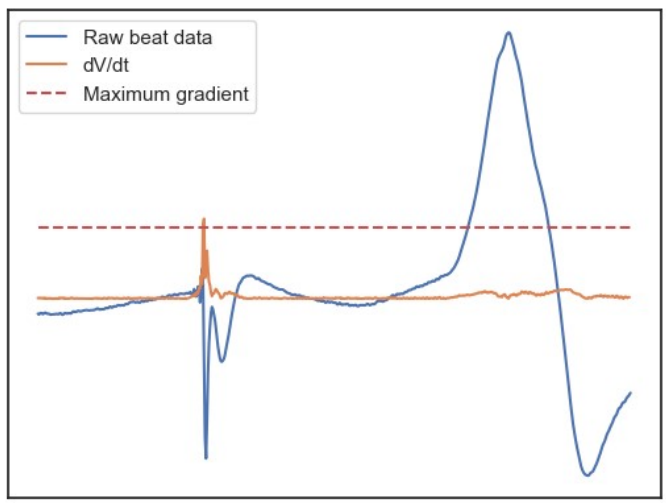


Figure 13: Example of excluded P-QRS interval. Pacing results in a gradient, $\frac{dV}{dt}$, above 0.15.

3.2.3 P-wave feature extraction

The P-wave feature extraction constitutes a fully automatic approach to P-wave delineation, using the wt-delineator library [63]. The 12 ECG-leads from the segments are fed into the `delineateMultiLeadECG(sig,fs)` function, returning indices for $P_{on}(lead)$ and $P_{end}(lead)$, which represent P-wave onset and offset. These indices are different for the different leads, and certain leads are used for certain features. Features are extracted following the methods defined in research [64]. Code for the feature extraction can be found in Appendix.

P-wave duration is calculated using the 6 frontal plane leads (I, II, III, aVR, aVL, aVF). An average index value, for each P-wave, across the frontal leads are used as the index for avg_P_{on} and avg_P_{end}. In some cases indices for P-waves are missing in the separate leads, these leads are excluded in calculations of avg_P_{on} and avg_P_{end}. P-wave duration is calculated in Equation 1.

$$p\_dur(beat, seg) = avg\_P_{end}(beat, seg) - avg\_P_{on}(beat, seg) \quad (1)$$

P-wave voltage is measured in lead I. The value is calculated by subtracting the minimum voltage (see Eq. 3) from the maximum voltage (see Eq. 4) measured between the avg_P_{on} and avg_P_{end} index, as described in Equation 2.

$$p\_volt(beat, seg) = max\_volt(beat, seg) - min\_volt(beat, seg) \quad (2)$$

where,

$$min\_volt\left(\{f(x) : x = avg\_P_{on}, ..., avg\_P_{end}\}\right), \quad (3)$$

$$max\_volt\left(\{f(x) : x = avg\_P_{on}, ..., avg\_P_{end}\}\right) \quad (4)$$

P-wave area is calculated using the previously extracted duration and voltage features, as described in Equation 5:

$$p\_area(beat, seg) = \frac{1}{2} p\_dur(beat, seg) \times p\_volt(beat, seg) \quad (5)$$

Feature Statistical Analysis is visualized using a seaborn `pairplot`, showing relationships between features and distribution of the labels for all features.

3.3 Machine Learning

3.3.1 Clinical data

Data is labeled as AF recurrent if there is a recorded AF/atrial tachycardia (AT)/atrial flutter (AFL) within 12 months postablation. Data is labeled as non AF recurrent if non of the mentioned arrhythmias occur within 12 months postablation. Only patients with a 12 month follow-up are included, therefore 29 patients are excluded. The three datasets are concatenated to ease data-handling, beats for each patient are still kept separate. The train-test-split is done patient-wise with a 80-20 split. Training for the RF and CNNs, was accomplished on a single Intel(R) Iris(R) Xe Graphics GPU.

Patient balancing is done as there was an imbalance between beats per patient in the data after beat classification. In order to combat a patient bias in training data, every patient was sampled by 100 beats. Patients with more than 100 beats are under-sampled and patients with less than 100 beats are over-sampled, beats are over-sampled at random from the available usable beats.

Balancing the dataset is done to combat bias of the majority label. Balancing was done by undersampling the majority label. To keep a wide array of patients each patient has a number of beats removed resulting in a $50/50$ split of the labels.

3.3.2 A Synthesized Dataset

In order to evaluate the ability of the CNN to extract information from P-wave morphology, a separate dataset is created. The ECG data for subjects labeled as 'AF recurrent' is used for synthetization of P-wave attributes. For each 'AF recurrent' subject, the two first P-QRS intervals are saved. Each P-QRS interval is analysed and the center, in between P-wave onset and offset, is annotated manually by an operator (see Figure 14). P-waves are manipulated using a Hanning window function, defined in Eq. 6. For each P-QRS interval, a random Hanning window is generated based on two parameters, amplitude (see Eq. 7) and width (see Eq. 8) Randomly generated Hanning window functions can be seen in Figure 15. The Hanning window is then multiplied to the P-QRS intervals, generating more emphasized P-waves.

An example of the synthetization can be seen in Figure 16. To achieve a balanced dataset, the same amount of P-QRS intervals for subjects labeled as 'non AF recurrent', are added, resulting in a $50/50$ split. The train-test-split is done patient-wise using 10-fold cross-validation.

$$w(n) = a\left(1.5 - 0.5cos\left(\frac{2\pi n}{M-1}\right)\right) \quad 0 \leq n \leq M-1 \tag{6}$$

Where,

$$a \in [1.3, 2.5] \text{ and } a \text{ is randomly chosen} \tag{7}$$

$$M \in [100ms, 160ms] \text{ and } M \text{ is randomly chosen} \tag{8}$$

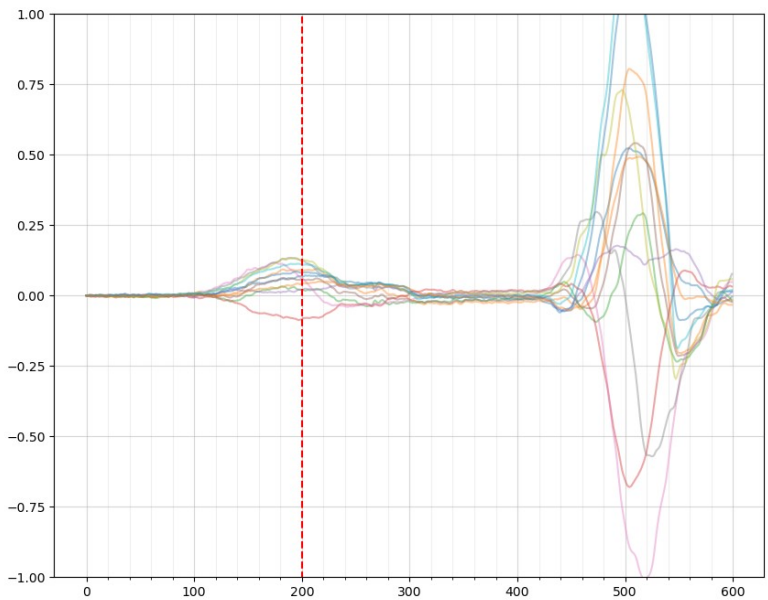


Figure 14: 12-lead, P-QRS interval annotated with red dashed line at P-wave center.

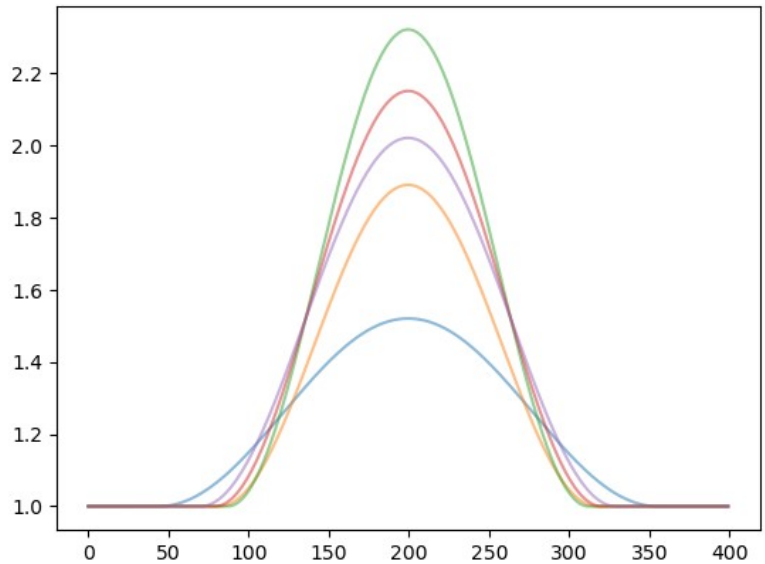


Figure 15: 5 Hanning window functions with randomly generated parameters for amplitude and width.

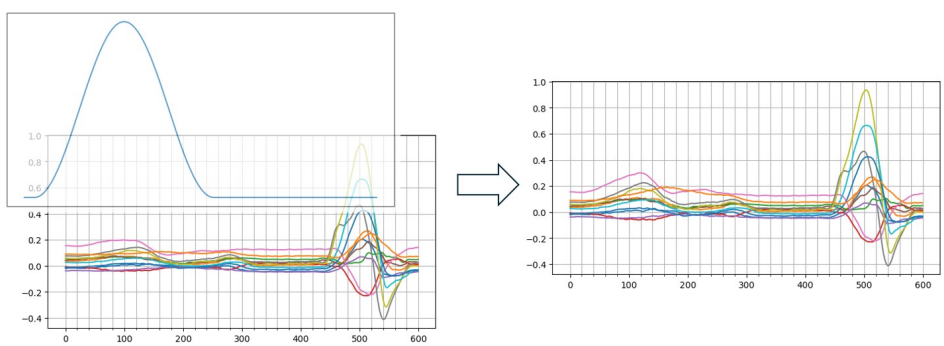


Figure 16: P-wave synthetization using a Hanning window function.

3.3.3 Random forest Classifier

The Random Forest Classifier was setup using the sklearn library. It was implemented with a max depth of 3 and fitted to the extracted P-wave features.

3.3.4 Convolutional Neural Network

A 4-layer CNN structure is used, comprised of: input, 1D convolution, pooling, fully connected, and output-layer. The input shape is one dimension of

600 samples for 17 channels. A summary of the architecture is provided in Table 2. For the first convolutional layer, a kernel width of 300 samples, 150 ms, is used. With this width, the kernel is able to cover the P-wave, which is the area of interest. Training parameters were set to a batch normalization of 32 and 10 epoch.

Layer	Type	Filters	Kernel size	Activation
1	Convolution	600 x 16	1 x 300	relu
2	Max-pooling	300 x 16	1 x 3	-
3	Convolution	300 x 64	1 x 100	relu
4	Max-pooling	100 x 64	1 x 3	-
5	Fully-connected	1 x 256	-	relu
6	Fully-connected	1 x 1	-	sigmoid

Table 2: CNN architecture.

3.4 Model Evaluation

3.4.1 Evaluation Metrics

The performance of the models were measured using `sklearn.metric`, functions `f1_score`, `recall_score`, `precision_score`, `roc_auc_score` and `accuracy_score`. Abbreviations are as follows: True Positive (TP, i.e. 'non AF recurrent' classified as 'non AF recurrent'), False Positive (FP i.e. 'AF recurrent' classified as 'non AF recurrent'), True Negative (TN, i.e. 'AF recurrent' classified as 'AF recurrent'), False Negative (FN i.e. 'non AF recurrent' classified as 'AF recurrent').

Recall, Precision, F1-score and Accuracy are calculated as follows:

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

The ROC AUC score is the area under the ROC curve. The ROC curve shows possible rates of TP to FP and visualizes the classifiers performance at these different rates.

A confusion matrix is used to display classifier performance. Each row represents the true label and each column represents the predicted label.

3.4.2 k-Fold Cross-Validation

A 10-fold cross-validation is used to validate the CNN trained on synthesized data.

4 Results

4.1 Pre-processing

4.1.1 Visual Assessment

The visual assessment resulted in 230 recordings deemed usable for further analysis, out of 281 recordings. Out of the 281 recordings, 149 were annotated to have some kind of pacing stimulus within the segment. Usable recordings are defined as recordings with 60s segments of sufficient quality.

Dataset	Available recordings	Usable recordings	<60s recordings
Dataset 1	39	32	8
Dataset 2	58	54	10
Dataset 3	184	144	0
Total	281	230	18

Table 3: Resulting recordings post visual assessment.

4.1.2 Beat classification

After beat classification there were 23 recordings excluded since all of the P-QRS intervals were excluded based on the criteria in Sec. 3.2.2, leaving 207 recordings for further analysis. Dataset 3 had more P-QRS intervals removed than Dataset 1 and 2, as seen in Figure 17. Rejection statistics, Table 4, show a large number of P-QRS intervals removed due to short RR-interval(high BPM), followed by correlation and gradient exclusion removing a small number of P-QRS intervals.

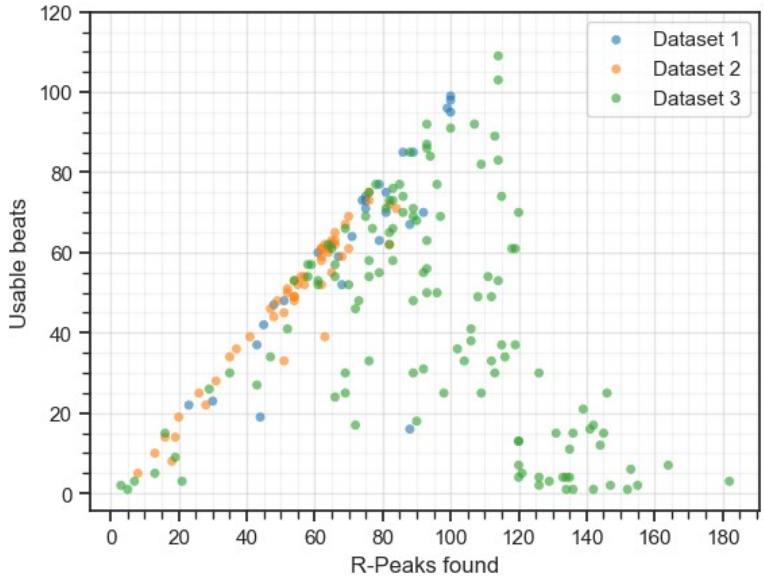


Figure 17: P-QRS intervals retained plotted over detected R-peaks found in recordings. Recordings where 100% of P-QRS intervals are retained follow a line starting from the origin with a gradient of 1.

Dataset	RR-interval (%, range)	Gradient (%, range)	Correlation (%, range)
Dataset 1	8%, (1-67)	<1%, (0-4)	4.6%, (0-20)
Dataset 2	4.4%, (1-16)	<1%, (0-8)	4.6%, (0-13)
Dataset 3	46.5%,(1-179)	<1%, (0-4)	3.4%, (0-20)
Total	19.6%	<1%	4.2%

Table 4: Average and range of P-QRS intervals removed for each filtered recording, per dataset.

Dataset	P-QRS intervals retained (mean, range)
Dataset 1	63, (19-99)
Dataset 2	48, (5-75)
Dataset 3	46, (1-114)
Total	52

Table 5: Average and range of P-QRS intervals retained for each filtered recording, per dataset.

4.1.3 Statistical Analysis of Features

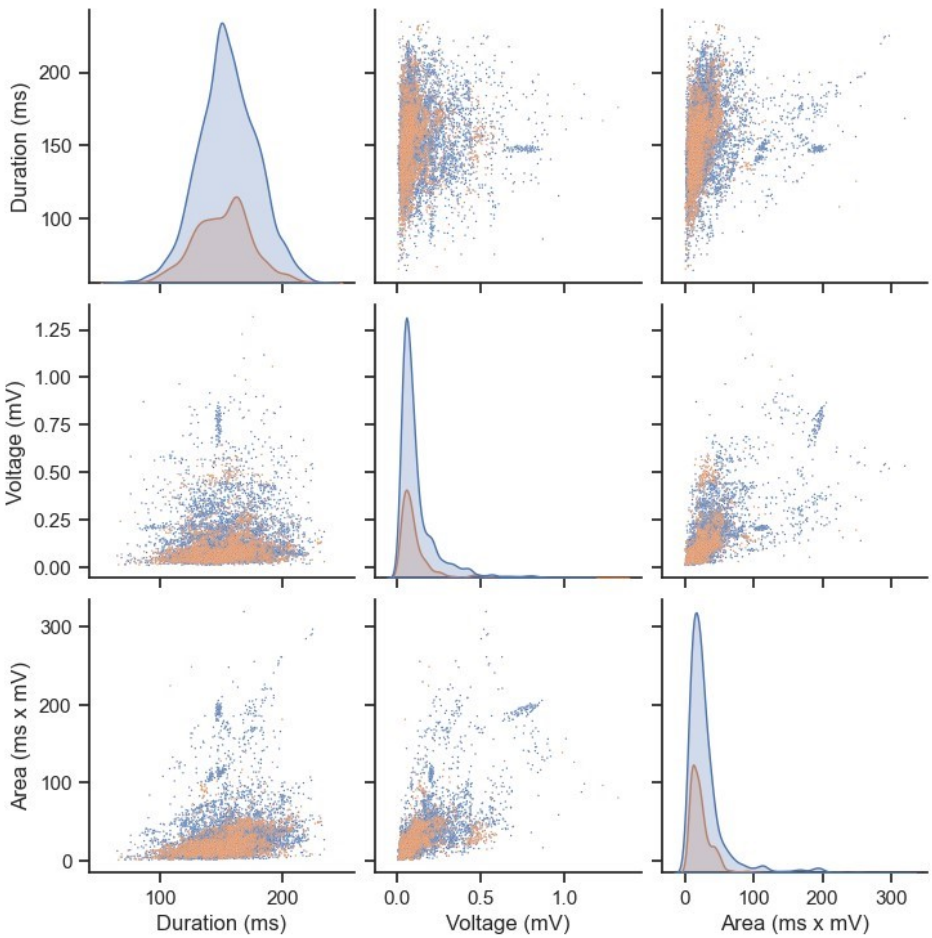


Figure 18: Plot of extracted P-wave features. Recurring and non-recurring AF with features *p_dur*, *p_volt* and *p_area*. The diagonal plots show the distribution of the classes for the features. Non-recurring AF in blue and recurring AF in orange.

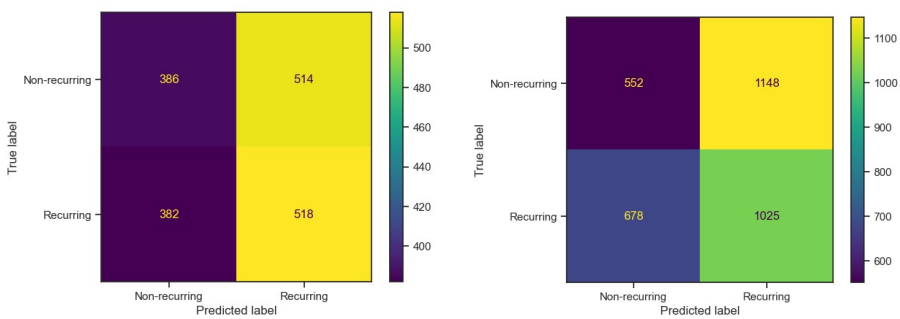
As illustrated in Figure 18, the visualization of the features show a similar distribution, for all features, for both classes. Some outliers can be seen for all features and classes.

4.2 Prediction of AF recurrence

Table 6 shows the performance metrics for two models, a RF classifier trained on extracted P-wave features and a CNN trained on raw ECG and CS catheter electrogram data. Figure 19 illustrates two confusion matrices with predictions and true labels for the two models.

Models	F1	Recall	Precision	AUC	Accuracy
RF P-wave features	0.536	0.576	0.502	0.502	0.502
CNN Raw data	0.528	0.602	0.472	0.463	0.463

Table 6: Model Performance Metrics for AF recurrence prediction.



(a) Random Forest Classifier (b) CNN Raw Data

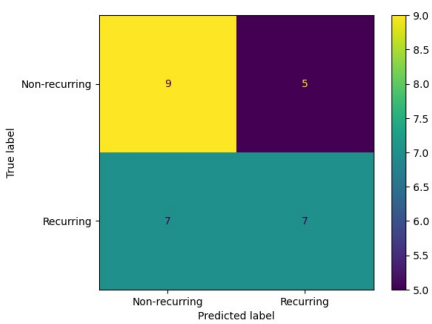
Figure 19: Confusion matrices for both models.

4.3 CNN Model Evaluation

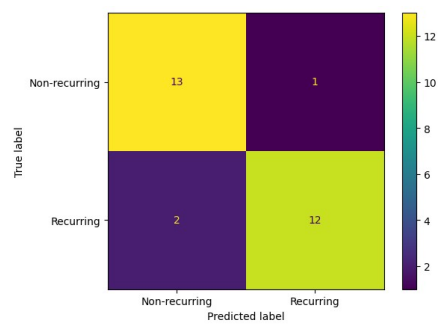
The generated synthesized dataset included 276 P-QRS intervals, 138 'AF recurrent' and 138 'non-AF recurrent'. Performance metrics, using 10-fold cross-validation, for a CNN trained on synthesized ECG data are shown in Table 7, with an increase in all metrics compared to previous CNN. Figure 20 illustrates the confusion matrices for the most and least accurate fold in cross-validation, as well as the summary of predictions for all folds.

Model	F1	Recall	Precision	AUC	Accuracy
CNN Synthesized Data	0.794	0.797	0.791	0.793	0.793

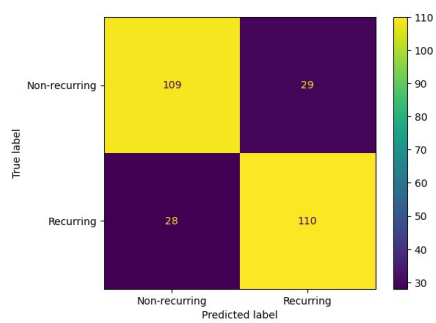
Table 7: Model Performance Metrics for CNN trained on synthesized data.



(a) Prediction for the least accurate model.



(b) Prediction for the most accurate model.



(c) Summary of predictions from all models in 10-fold cross-validation.

Figure 20: Confusion matrices for CNN trained on synthesized data.

5 Discussion

In this study P-wave features, ECG and CS catheter electrogram data is extracted from the present dataset of 281 subjects, using automatic delineation. A Random Forest (RF) classifier is trained on the extracted P-wave features and a CNN is trained on ECG and CS catheter electrogram data to predict 12-month AF recurrence post-PVI ablation. For both the RF and CNN classifier, F1-score and accuracy are around 0.5 (see Table 6). Similarly, the confusion matrices show a random distribution of predictions (Figure 19 a, b). The AUC-score for both models are also around 0.5, which is similar to tossing a coin [65]. Therefore the models have no capability of predicting AF recurrence on this dataset. Moreover, the CNN is evaluated by creating a synthesized dataset, inserting augmented P-wave characteristics in raw ECG-data. The model accuracy and AUC-score is 0.793 (see Table 7), illustrating the CNN capability of extracting information similar to that of P-wave morphology from an ECG. The predictions in the confusion matrices, in Figure 20, show an even distribution between FP and FN, with the worst model of the cross-validation not being able to predict AF recurrence in the subset testdata.

There is no denying the overall potential of CNNs finding patterns in data. CNNs used on medical images find complex pathological patterns to aid in diagnosis [66]. Therefore, it is hypothesised that a 1-D CNN, similarly, is capable of finding complex information in 1-D data. The complexity of the augmented P-wave characteristics lie in that each applied Hanning window is generated from random values of amplitude and width, and that this is applied at the annotated P-wave center, which is different for each P-QRS interval. Since the Hanning window is multiplied to the existing raw ECG, the original morphology is not lost, but amplified. This amplification can be as much as 2.5 times, which is noticeably different compared to the non-synthesized P-QRS intervals. The method of evaluating the CNN using synthesized dataset concludes that the CNN can extract valuable information from ECG data regardless of the P-wave location, although the amplification of the P-wave in the synthesized dataset is far beyond the small differences in P-wave morphology used in diagnosis of atrial health [67].

5.1 Methods

The visual assessment is difficult to evaluate as there are no labels for adequate signal quality, for the recordings. The CS catheter leads (see Figure 9(C)) are good indicators of AF as it clearly records arrhythmic activation in the atria, deemed to be of inadequate signal quality (see Figure 10(B)). Pacing was present in 149 out of all recordings. How many of these recordings that were included in further processing was not recorded. The decision to include these recordings was made to allow for a larger dataset to be used in training. The inclusion also made training data more realistic, as the overall prevalence of pacing in recordings was high. Valuable information of atrial health can still be obtained from paced P-wave morphology [68]. In a study, Zvuloni et al [69] perform segment extraction on 12-lead ECG data for AF recurrence prediction. A signal quality criterion was based on a moving signal quality index (bSQI), with 18 out of 130 recordings being excluded because of low quality. This is done for the same purpose as for the visual assessment and beat rejection, to extract data of good signal quality. As mentioned, there are no labels of the recordings for this present study, so there can be no performance comparison. A study evaluating the performance of bSQI saw an accuracy of 92.5%, 95.9% sensitivity and 90.1% specificity [70] in classifying ECG segment quality. Applying this method to the present study would give a baseline number of ECG signal quality, while also eliminating induced operator bias in the visual assessment, as even expert operators show variance in labelling data [71].

Similarly, the method of beat classification cannot be thoroughly evaluated as there are no labels of signal quality for the ECG signals. Neither were the applied criteria evaluated, proving it difficult to discard the hypothesis of inadequate P-QRS intervals being included in training. Results of P-QRS interval rejection in Table 4, reveal that the RR-interval criteria removes the majority of the P-QRS interval, which could explain the low number of P-QRS intervals removed for the following criterias as the RR-interval criteria is applied first. The cutoff value for the RR-interval criteria is set quite high at >120BPM. It is worth mentioning that tachycardia is defined as >100BPM [72]. This could perhaps cause T-wave influence on the P-wave morphology. A similar decision to the one made on inclusion of pacing was taken with regards to the cutoff for the RR-interval, as a lower value would remove vast amounts of data. Furthermore, the results of the gradient and

correlation criteria, summarized in Table 4, show a low impact across the datasets. Given the inclusion of recordings with pacing, it was hypothesized that the gradient criteria would have a more significant impact. Further evaluation of these criterias is therefore necessary.

Martinez et al [22] reports, in a paper written on the performance of the wt-delineator, a mean and standard deviation error for P-wave onset and P-wave offset of $2 \pm 14.8ms$ and $1.9 \pm 12.8ms$, which could lead to inaccurate calculations of *p_dur* in the proposed method. wt-delineator was used for automatic delineation, annotating *P_on*, *P_end* and R-peak, delineation error could contribute to a poor calculation of *p_dur* and *p_volt* [63]. As illustrated in Figure 18, the statistical analysis shows no sign of a cutoff value in *p_dur* between AF recurring an non-recurring patients. Neither is there a clear separation in the *p_volt* for a cutoff value at 0.16mV of that found by Nakatani et al [21]. *p_volt* is also highly impacted by the accuracy of *P_on* and *P_end*. This is also implied by the RF classifier, as it is not able to conform to a cutoff value for the features to allow for good a classification (see Figure 19(a)).

Balancing of the dataset based on beats per patient was implemented using oversampling. This was applied to both the extracted P-wave features and the ECG data. The usable beats were copied until each patient had 100 beats. Patients with a low amount of usable beats therefore had their usable beats over-sampled to a larger extent. There are numerous patients with a low amount of usable beats, shown in the bottom of Figure 17. The presence of multiples of beats could impact training of the model, creating a bias in prediction.

5.2 Data

Figure17 and Table 4 illustrates the inhomogeneity of the datasets. This could hint towards differences when recording ECG, use of different EP systems, different treatments, or demographic differences between the populations. For prediction models, development and validation with the use of different populations legitimizes the performance of the model across populations [73]. However, possible differences in these datasets, coupled with the small amount of patients in each dataset, might be a factor contributing

to the lower prediction accuracy [74]. Another study, with similar aims, used a dataset including more than 100 000 subjects [75].

Studies by Jadidi et al [17] and Chen et al [76] use recordings before ablation treatment only. In the present study, the datasets contained recordings from pre-, post- and during ablation treatment, which could make it more difficult for the model to fit to the data, as research has shown that P-wave characteristics change with ablation [77].

There was an error in generating the synthesized dataset for evaluating the CNN performance, the Hanning window applied to the signal widened the duration, which follows research findings [17], but increased P-wave amplitude which goes against the patophysiology of AF recurrence P-wave characteristics [21]. However, the aim of using the synthesized dataset is to evaluate if the CNN is capable of extracting information in P-wave characteristics, therefore it does not matter if the synthetization creates patophysiologically accurate changes. Albeit, a more patophysiologically accurate dataset would allow for a robust evaluation of the CNN.

The scores of the CNN training on synthesized data (see Table 7) illustrate a capability of the CNN to extract information from the synthesized ECG. Thus, it may be be argued that the raw ECG of the dataset does not contain distinguishable abnormalities in P-wave morphology, which would cause the low scores of the CNN.

With all the limitations in this methodology and dataset, the limited performance of the models can be linked to the principle of "poor quality in - poor quality out", which recognizes that poor quality data fed into the models creates an unreliable output [78]. Furthermore, this allows us to not disqualify the potential of the proposed models in classification of AF recurrence.

5.3 Survey of the field

A more successful attempt at developing a CNN to predict AF ablation outcomes is the proposed method by Tang et al [52], similarly developing a 1D-CNN using ECG and iECG data, but also leveraging clinical features to predict outcome. With a dataset including 156 patients the resulting AUC-

score is 0.859 ± 0.082 using ECG, iECG and clinical data, and 0.833 ± 0.084 using only ECG and iECG. Similarly Attia et al [75] developed a 1D-CNN classifying presence of AF with an AUC-score of 0.87. Both studies used a time series as input for training of 5-sec and 10-sec, respectively. This allows the CNN to extract features across multiple beats.

Another study investigating AF recurrence using ML is Zvuloni et al [69]. In their study, two RF classifiers leveraged features extracted from 60-sec 12-lead ECG segments from before and after ablation to classify AF recurrence, resulting in AUC-scores of 0.56 for pre-ablation and 0.67 for post-ablation data. Manually crafted features were based on heart rate variability and morphology.

The RF classifier uses the feature *p_dur* to no apparent success. Contrary to the present study, in research by Jadidi et al [23], similar features are extracted and used with a neural network achieving great results. The neural network, trained on SR PWD and morphology, is able to diagnose left ACM with an AUC-score of 0.85. Furthermore, this classification can be used to determine rates of AF recurrence, where patients with left ACM had a higher amount of arrhythmia recurrence. In other work by Jadidi et al [79], 95 subjects were used for classifiyng the extent of LA-LVS in patients, based on P-wave morphology and PWD. Patients were classified into three groups dependent on the extent of LA-LVS, with freedom from arrhythmia for the separate groups of 77%, 53%, and 33%.

5.4 Future work

As the aforementioned studies reiterates, the P-wave morphology carries valuable information about atrial health and rate of absence of AF post ablation [17]. Moreover, a ML model, particularly the CNN approach to extracting information from ECG signals, exhibit potential [52]. To fully investigate the potential of the method proposed in this study, the pre-processing of data needs further work. For a more robust and standardized signal quality, it would be interesting to implement quality analysis using SQI instead of a manual, visual assessment of the data. Furthermore with the need for intricate delineation of the P-wave, the automatic delineation should be validated

using expert annotations. And as always with ML in healthcare, a big and diverse dataset should be utilized to ensure robust and accurate model performance. With the increase of AF, the development of an automatic solution to assert patient risk for AF recurrence post-PVI is of great importance. Hence, more research on the subject will be needed.

5.5 Conclusion

This work investigates the predictive capabilities of two models trained on P-wave features and ECG and CS catheter electrogram, respectively. A preprocessing method based on automatic delineation is proposed, eliminating the need for extensive manual annotation of ECG data. Both models show no sign of predicting AF recurrence on the present dataset, with AUC-scores of 0.502 and 0.463. Furthermore, through training and testing on synthesized data, and a AUC-score of 0.793, the CNN exhibits potential in extracting information from P-wave morphology. Moreover, evaluation of the automatic delineation is needed to rule out annotation errors affecting training data. Future work should focus on implementing an automatic delineation coupled with automatic signal quality analysis and apply the method to a larger dataset.

References

[1] Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. European Heart Journal. 2020 08;42(5):373-498. Available from: https://doi.org/10.1093/eurheartj/ehaa612.

[2] Zeitler EP, Ronk CJ, Cockerham A, Huse S, McKindley DS, Kim MH. Healthcare resource utilization in patients with newly diagnosed atrial fibrillation in the United States. Expert Review of Pharmacoeconomics & Outcomes Research. 2022;22(5):763-71. PMID: 35209794. Available from: https://doi.org/10.1080/14737167.2022.2045955.

[3] Jiang S, Seslar SP, Sloan LA, Hansen RN. Health care resource utilization and costs associated with atrial fibrillation and rural-urban disparities. Journal of Managed Care & Specialty Pharmacy. 2022;28(11):1321-30. PMID: 36282926. Available from: https://doi.org/10.18553/jmcp.2022.28.11.1321.

[4] Pappone C, Augello G, Sala S, Gugliotta F, Vicedomini G, Gulletta S, et al. A Randomized Trial of Circumferential Pulmonary Vein Ablation Versus Antiarrhythmic Drug Therapy in Paroxysmal Atrial Fibrillation. Journal of the American College of Cardiology. 2006;48(11):2340-7. Available from: https://www.jacc.org/doi/abs/10.1016/j.jacc.2006.08.037.

[5] Wazni OM, Marrouche NF, Martin DO, Verma A, Bhargava M, Saliba W, et al. Radiofrequency Ablation vs Antiarrhythmic Drugs as First-line Treatment of Symptomatic Atrial FibrillationA Randomized Trial. JAMA. 2005;293(21):2634-40. Available from: https://doi.org/10.1001/jama.293.21.2634.

[6] Blomström-Lundqvist C, Gizurarson S, Schwieler J, Jensen SM, Bergfeldt L, Kennebäck G, et al. Effect of Catheter Ablation vs Antiarrhythmic Medication on Quality of Life in Patients With Atrial

Fibrillation: The CAPTAF Randomized Clinical Trial. JAMA. 2019 03;321(11):1059-68. Available from: https://doi.org/10.1001/jama.2019.0335.

[7] Tilz RR, Heeger CH, Wick A, Saguner AM, Metzner A, Rillig A, et al. Ten-Year Clinical Outcome After Circumferential Pulmonary Vein Isolation Utilizing the Hamburg Approach in Patients With Symptomatic Drug-Refractory Paroxysmal Atrial Fibrillation. Circulation: Arrhythmia and Electrophysiology. 2018;11(2):e005250. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCEP.117.005250.

[8] Poole JE, Bahnson TD, Monahan KH, Johnson G, Rostami H, Silverstein AP, et al. Recurrence of Atrial Fibrillation After Catheter Ablation or Antiarrhythmic Drug Therapy in the CABANA Trial. Journal of the American College of Cardiology. 2020;75(25):3105-18. Available from: https://www.sciencedirect.com/science/article/pii/S0735109720351731.

[9] Tzeis S, Gerstenfeld EP, Kalman J, et al. European Heart Rhythm Association/Heart Rhythm Society/Asia Pacific Heart Rhythm Society/Latin American Heart Rhythm Society expert consensus statement on catheter and surgical ablation of atrial fibrillation. Journal of Interventional Cardiac Electrophysiology. 2024.

[10] Jadidi AS, Lehrmann H, Keyl C, Sorrel J, Markstein V, Minners J, et al. Ablation of Persistent Atrial Fibrillation Targeting Low-Voltage Areas With Selective Activation Characteristics. Circulation: Arrhythmia and Electrophysiology. 2016;9(3):e002962. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCEP.115.002962.

[11] Saglietto A, Gaita F, Blomstrom-Lundqvist C, Arbelo E, Dagres N, Brugada J, et al. AFA-Recur: an ESC EORP AFA-LT registry machine-learning web calculator predicting atrial fibrillation recurrence after ablation. EP Europace. 2022;25(1):92-100. Available from: https://doi.org/10.1093/europace/euac145.

[12] D'Ascenzo F, Corleto A, Biondi-Zoccai G, Anselmino M, Ferraris F, di Biase L, et al. Which are the most reliable predictors of recurrence of atrial fibrillation after transcatheter ablation?: a meta-analysis. International Journal of Cardiology. 2013;167(5):1984-

9. Available from: https://www.sciencedirect.com/science/article/pii/S0167527312006195.

[13] Yang G, Yang B, Wei Y, Zhang F, Ju W, Chen H, et al. Catheter Ablation of Nonparoxysmal Atrial Fibrillation Using Electrophysiologically Guided Substrate Modification During Sinus Rhythm After Pulmonary Vein Isolation. Circulation: Arrhythmia and Electrophysiology. 2016;9(2):e003382. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCEP.115.003382.

[14] Yagishita A, De Oliveira S, Cakulev I, et al. Correlation of left atrial voltage distribution between sinus rhythm and atrial fibrillation: identifying structural remodeling by 3-D electroanatomic mapping irrespective of the rhythm. Journal of Cardiovascular Electrophysiology. 2016;27:905-12.

[15] Rolf S, Kircher S, Arya A, Eitel C, Sommer P, Richter S, et al. Tailored Atrial Substrate Modification Based on Low-Voltage Areas in Catheter Ablation of Atrial Fibrillation. Circulation: Arrhythmia and Electrophysiology. 2014;7(5):825-33. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCEP.113.001251.

[16] Seemann F, Baldassarre LA, Llanos-Chea F, Gonzales RA, Grunseich K, Hu C, et al. Assessment of diastolic function and atrial remodeling by MRI – validation and correlation with echocardiography and filling pressure. Physiological Reports. 2018;6(17):e13828. Available from: https://physoc.onlinelibrary.wiley.com/doi/abs/10.14814/phy2.13828.

[17] Jadidi A, Müller-Edenborn B, Chen J, Keyl C, Weber R, Allgeier J, et al. The Duration of the Amplified Sinus-P-Wave Identifies Presence of Left Atrial Low Voltage Substrate and Predicts Outcome After Pulmonary Vein Isolation in Patients With Persistent Atrial Fibrillation. JACC: Clinical Electrophysiology. 2018;4(4):531-43. Available from: https://www.sciencedirect.com/science/article/pii/S2405500X17311866.

[18] Pranata R, Yonas E, Vania R. Prolonged P-wave duration in sinus rhythm pre-ablation is associated with atrial fibrillation recurrence after pulmonary vein isolation–A systematic review and meta-analysis. Annals of Noninvasive Electrocardiology. 2019;24(5):e12653. Available

from: https://onlinelibrary.wiley.com/doi/abs/10.1111/anec.12653.

[19] Wu JT, Long DY, Dong JZ, Wang SL, Fan XW, Yang HT, et al. Advanced interatrial block predicts clinical recurrence of atrial fibrillation after catheter ablation. Journal of Cardiology. 2016;68(4):352-6. Frontiers of Echocardiography. Available from: https://www.sciencedirect.com/science/article/pii/S0914508715003573.

[20] Saclova L, Nemcova A, Smisek R, et al. Reliable P wave detection in pathological ECG signals. Scientific Reports. 2022;12:6589. Available from: https://doi.org/10.1038/s41598-022-10656-4.

[21] Nakatani Y, Sakamoto T, Yamaguchi Y, Tsujino Y, Kataoka N, Kinugawa K. P-wave vector magnitude predicts recurrence of atrial fibrillation after catheter ablation in patients with persistent atrial fibrillation. Annals of Noninvasive Electrocardiology. 2019;24(5):e12646. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/anec.12646.

[22] Martinez JP, Almeida R, Olmos S, Rocha AP, Laguna P. A wavelet-based ECG delineator: evaluation on standard databases. IEEE Transactions on Biomedical Engineering. 2004;51(4):570-81.

[23] Jadidi A, Pilia N, Huang T, Mueller-Edenborn B, Allgeier J, Nairn D, et al. Comparison of amplified P-wave analysis to artificial intelligence-derived analysis for diagnosis of atrial cardiomyopathy and outcome prediction following pvi for persistent atrial fibrillation. EP Europace. 2023;25(Supplement_1):euad122.531. Available from: https://doi.org/10.1093/europace/euad122.531.

[24] Hsieh CH, Li YS, Hwang BJ, Hsiao CH. Detection of Atrial Fibrillation Using 1D Convolutional Neural Network. Sensors. 2020;20(7). Available from: https://www.mdpi.com/1424-8220/20/7/2136.

[25] Betts J, Desaix P, Johnson E, Johnson J, Korol O, Kruse D, et al . Anatomy and Physiology. OpenStax College; 2013.

[26] Barnett VA. In: Iaizzo PA, editor. Cellular Myocytes. Cham: Springer International Publishing; 2015. p. 201-14. Available from: https://doi.org/10.1007/978-3-319-19464-6_12.

[27] Laske TG, Shrivastav M, Iaizzo PA. In: Iaizzo PA, editor. The Cardiac Conduction System. Cham: Springer International Publishing; 2015. p. 215-33. Available from: https://doi.org/10.1007/978-3-319-19464-6_13.

[28] Wikimedia Commons. Heart; conduction system; 2007. [Online; accessed May 18, 2024]. Available from: https://en.wikipedia.org/wiki/File:Cardiac_Conduction_System.jpg.

[29] Vieau S, Iaizzo PA. In: Iaizzo PA, editor. Basic ECG Theory, 12-Lead Recordings, and Their Interpretation. Cham: Springer International Publishing; 2015. p. 321-34. Available from: https://doi.org/10.1007/978-3-319-19464-6_19.

[30] Meek S, Morris F. ABC of clinical electrocardiography: Introduction. I—Leads, rate, rhythm, and cardiac axis. BMJ (Clinical research ed). 2002 03;324:415-8.

[31] Wikimedia Commons. Spatial orientation of EKG leads; 2015. [Online; accessed May 18, 2024]. Available from: https://commons.wikimedia.org/wiki/File:EKG_leads.png.

[32] Wikimedia Commons. Schematic diagram of normal sinus rhythm for a human heart as seen on ECG.; 2007. [Online; accessed May 18, 2024]. Available from: https://commons.wikimedia.org/wiki/File:SinusRhythmLabels.svg.

[33] Issa ZF, Miller JM, Zipes DP. 4 - Electrophysiological Testing: Tools and Techniques. In: Issa ZF, Miller JM, Zipes DP, editors. Clinical Arrhythmology and Electrophysiology (Third Edition). third edition ed. Philadelphia: Elsevier; 2019. p. 81-124. Available from: https://www.sciencedirect.com/science/article/pii/B9780323523561000049.

[34] Karthan M, Martin R, Holl F, Swoboda W, Kestler H, Pryss R, et al. Enhancing mHealth data collection applications with sensing capabilities. Frontiers in Public Health. 2022 09;10.

[35] Li N, He F, Ma W, Wang R, Jiang L, Zhang X. The Identification of ECG Signals Using Wavelet Transform and WOA-PNN. Sensors. 2022;22(12). Available from: https://www.mdpi.com/1424-8220/22/12/4343.

[36] Lenis G, Pilia N, Oesterlein T, Luik A, Schmitt C, Dössel O. P wave detection and delineation in the ECG based on the phase free stationary wavelet transform and using intracardiac atrial electrograms as reference. Biomedical Engineering / Biomedizinische Technik. 2016;61(1):37-56. Available from: https://doi.org/10.1515/bmt-2014-0161 [cited 2024-05-19].

[37] Huang JC, Wei SY, Chen SC, Chang JM, Hung CC, Su HM, et al. P Wave Dispersion and Maximum P Wave Duration Are Associated with Renal Outcomes in Chronic Kidney Disease. PLoS ONE. 2014;9(7):1 7. Available from: https://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=a9h&AN=97361055&site=eds-live&scope=site.

[38] Benjamin EJ, Chen PS, Bild DE, Mascette AM, Albert CM, Alonso A, et al. Prevention of Atrial Fibrillation. Circulation. 2009;119(4):606-18. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.108.825380.

[39] Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. European Heart Journal. 2020 08;42(5):373-498. Available from: https://doi.org/10.1093/eurheartj/ehaa612.

[40] Joglar JA, Chung MK, Armbruster AL, Benjamin EJ, Chyou JY, Cronin EM, et al. 2023 ACC/AHA/ACCP/HRS Guideline for the Diagnosis and Management of Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation. 2024;149(1):e1-e156. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIR.0000000000001193.

[41] ki Iwasaki Y, Nishida K, Kato T, Nattel S. Atrial Fibrillation Pathophysiology. Circulation. 2011;124(20):2264-74. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.111.019893.

[42] Verma A, Macle L, Cox J, Skanes AC. Canadian Cardiovascular Society Atrial Fibrillation Guidelines 2010: Catheter Ablation for Atrial Fibrillation/Atrial Flutter. Canadian Journal of Cardiology. 2011 Jan;27(1):60-6. Available from: https://doi.org/10.1016/j.cjca.2010.11.011.

[43] Hartl S, Makimoto H, Gerguri S, Clasen L, Kluge S, Brinkmeyer C, et al. Wide Antral Circumferential Re-Ablation for Recurrent Atrial Fibrillation after Prior Pulmonary Vein Isolation Guided by High-Density Mapping Increases Freedom from Atrial Arrhythmias. Journal of Clinical Medicine. 2023;12(15). Available from: https://www.mdpi.com/2077-0383/12/15/4982.

[44] O'Neill MD, Jaïs P, Takahashi Y, Jönsson A, Sacher F, Hocini M, et al. The stepwise ablation approach for chronic atrial fibrillation—Evidence for a cumulative effect. Journal of Interventional Cardiac Electrophysiology. 2006 Sep;16(3):153-67. Available from: https://doi.org/10.1007/s10840-006-9045-1.

[45] Darby AE. Recurrent Atrial Fibrillation After Catheter Ablation: Considerations For Repeat Ablation And Strategies To Optimize Success. Journal of atrial fibrillation. 2016;9 1:1427. Available from: https://api.semanticscholar.org/CorpusID:9371349.

[46] Ko JS, Kim SS, Jeong HK, Kim NH. Decision-making for recurrent atrial fibrillation after catheter ablation. Cardiovasc Prev Pharmacother. 2023;5(4):102-12. Available from: https://www.e-jcpp.org/journal/view.php?number=78.

[47] Baranchuk A, Bayés de Luna A. The P-wave morphology: what does it tell us? Herzschrittmachertherapie + Elektrophysiologie. 2015 Sep;26(3):192-9. Available from: https://doi.org/10.1007/s00399-015-0385-3.

[48] Magnani JW, Johnson VM, Sullivan LM, Gorodeski EZ, Schnabel RB, Lubitz SA, et al. P Wave Duration and Risk of Longitudinal Atrial Fibrillation in Persons 60 Years Old (from the Framingham Heart Study). The American Journal of Cardiology. 2011;107(6):917-21.e1. Available from: https://www.sciencedirect.com/science/article/pii/S0002914910024240.

[49] German DM, Kabir MM, Dewland TA, Henrikson CA, Tereshchenko LG. Atrial Fibrillation Predictors: Importance of the Electrocardiogram. Annals of Noninvasive Electrocardiology. 2016;21(1):20-9. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/anec.12321.

[50] Rasmussen MU, Kumarathurai P, Fabricius-Bjerre A, Larsen BS, Domínguez H, Davidsen U, et al. P-wave indices as predictors of atrial fibrillation. Annals of Noninvasive Electrocardiology. 2020;25(5):e12751. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/anec.12751.

[51] Chen Q, Mohanty S, Trivedi C, Gianni C, Della Rocca DG, Canpolat U, et al. Association between prolonged P wave duration and left atrial scarring in patients with paroxysmal atrial fibrillation. Journal of Cardiovascular Electrophysiology. 2019;30(10):1811 1818. Available from: https://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=a9h&AN=139054100&site=eds-live&scope=site.

[52] Tang S, Razeghi O, Kapoor R, Alhusseini MI, Fazal M, Rogers AJ, et al. Machine Learning–Enabled Multimodal Fusion of Intra-Atrial and Body Surface Signals in Prediction of Atrial Fibrillation Ablation Outcomes. Circulation: Arrhythmia and Electrophysiology. 2022;15(8):e010850. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCEP.122.010850.

[53] Alloghani M, Al-Jumeily Obe D, Mustafina J, Hussain A, Aljaaf A. In: A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science; 2020. p. 3-21.

[54] Jung A. Machine Learning. The Basics. Machine Learning: Foundations, Methodologies, and Applications. Springer Nature Singapore; 2022. Available from: https://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=cat02271a&AN=atoz.ebs30578841e&site=eds-live&scope=site.

[55] Emanet N. ECG beat classification by using discrete wavelet transform and Random Forest algorithm. In: 2009 Fifth International Conference

on Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control; 2009. p. 1-4.

[56] Zou J, Han Y, So SS. Overview of artificial neural networks. Methods in molecular biology (Clifton, NJ). 2008;458:15—23. Available from: https://doi.org/10.1007/978-1-60327-101-1_2.

[57] Aggarwal CC. In: An Introduction to Neural Networks. Cham: Springer International Publishing; 2018. p. 1-52. Available from: https://doi.org/10.1007/978-3-319-94463-0_1.

[58] Ahmed AA, Ali W, Abdullah TAA, Malebary SJ. Classifying Cardiac Arrhythmia from ECG Signal Using 1D CNN Deep Learning Model. Mathematics. 2023;11(3). Available from: https://www.mdpi.com/2227-7390/11/3/562.

[59] Kiranyaz S, Ince T, Gabbouj M. Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. IEEE Transactions on Biomedical Engineering. 2016;63(3):664-75.

[60] Andersen RS, Peimankar A, Puthusserypady S. A deep learning approach for real-time detection of atrial fibrillation. Expert Systems with Applications. 2019;115:465-73. Available from: https://www.sciencedirect.com/science/article/pii/S0957417418305190.

[61] GE Healthcare UK. CardioLab Recording Systems; Accessed: 2024-04-29. Available from: https://www.gehealthcare.co.uk/products/ep-recording/cardio-lab-recording-systems.

[62] Boston Scientific Corporation. LABSYSTEM™ PRO, EP Recording System; 2024. [Online; accessed April 29, 2024]. Available from: https://www.bostonscientific.com/en-EU/products/capital-equipment-diagnostic/labsystem-pro.html.

[63] Ledezma CA. WTdelineator; 2018. Available at https://github.com/caledezma/WTdelineator. GitHub repository.

[64] Chen LY, Ribeiro ALP, Platonov PG, Cygankiewicz I, Soliman EZ, Gorenek B, et al. P Wave Parameters and Indices: A Critical Appraisal of Clinical Utility, Challenges, and Future Research—A Consensus Document Endorsed by the International Society of Electrocardiology and the International Society for Holter and Noninvasive

Electrocardiology. Circulation: Arrhythmia and Electrophysiology. 2022;15(4):e010435. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCEP.121.010435.

[65] Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. Continuing Education in Anaesthesia Critical Care Pain. 2008 12;8(6):221-3. Available from: https://doi.org/10.1093/bjaceaccp/mkn041.

[66] Daugaard Jørgensen M, Antulov R, Hess S, Lysdahlgaard S. Convolutional neural network performance compared to radiologists in detecting intracranial hemorrhage from brain computed tomography: A systematic review and meta-analysis. European Journal of Radiology. 2022;146:110073. Available from: https://www.sciencedirect.com/science/article/pii/S0720048X21005544.

[67] Censi F, Corazza I, Reggiani E, Calcagnini G, Mattei E, Triventi M, et al. P-wave Variability and Atrial Fibrillation. Scientific Reports. 2016 May;6(1):26799. Available from: https://doi.org/10.1038/srep26799.

[68] Fitzgerald DM, Hawthorne HR, Crossley GH, Simmons TW, Haisty WK. P wave morphology during atrial pacing along the atrioventricular ring: ECG localization of the site of origin of retrograde atrial activation. Journal of Electrocardiology. 1996;29(1):1-10. Available from: https://www.sciencedirect.com/science/article/pii/S0022073696801051.

[69] Zvuloni E, Gendelman S, Mohanty S, Lewen J, Natale A, Behar JA. Atrial Fibrillation Recurrence Risk Prediction From 12-Lead ECG Recorded Pre- and Post-Ablation Procedure. In: 2022 Computing in Cardiology (CinC). vol. 498; 2022. p. 1-4.

[70] Behar J, Oster J, Li Q, Clifford GD. ECG Signal Quality During Arrhythmia and Its Application to False Alarm Reduction. IEEE Transactions on Biomedical Engineering. 2013;60(6):1660-6.

[71] Zhu T, Dunkley N, Behar J, Clifton D, Clifford G. Fusing Continuous-Valued Medical Labels Using a Bayesian Model. Annals of biomedical engineering. 2015 03;43.

[72] Gopinathannair R, Olshansky B. Management of tachycardia. F1000Prime Rep. 2015 May;7:60.

[73] Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. Journal of Clinical Epidemiology. 2015 Mar;68(3):279-89. Available from: https://doi.org/10.1016/j.jclinepi.2014.06.018.

[74] Akinrinmade AO, Adebile TM, Ezuma-Ebong C, Bolaji K, Ajufo A, Adigun AO, et al. Artificial Intelligence in Healthcare: Perception and Reality. Cureus. 2023 Sep;15(9):e45594.

[75] Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. The Lancet. 2019;394(10201):861-7. Available from: https://www.sciencedirect.com/science/article/pii/S0140673619317210.

[76] Chen Q, Mohanty S, Trivedi C, Gianni C, Della Rocca DG, Canpolat U, et al. Association between prolonged P wave duration and left atrial scarring in patients with paroxysmal atrial fibrillation. Journal of Cardiovascular Electrophysiology. 2019;30(10):1811 1818. Available from: https://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=ccm&AN=139054100&site=eds-live&scope=site.

[77] Ortigosa N, Cano Sandberg F. Characterization of Changes in P-Wave VCG Loops Following Pulmonary-Vein Isolation. Sensors. 2021;21(5). Available from: https://www.mdpi.com/1424-8220/21/5/1923.

[78] Kilkenny MF, Robinson KM. Data quality: "Garbage in – garbage out". Health Information Management Journal. 2018;47(3):103-5. PMID: 29719995. Available from: https://doi.org/10.1177/1833358318774357.

[79] Müller-Edenborn B, Chen J, Allgeier J, Didenko M, Moreno-Weidmann Z, Neumann FJ, et al. Amplified sinus-P-wave reveals localization and extent of left atrial low-voltage substrate: implications for arrhythmia freedom following pulmonary vein isolation. EP Europace.

2019 11;22(2):240-9. Available from: https://doi.org/10.1093/europace/euz297.

Appendix

Figure 21: Recording with noisy leads.

PWD

```
# Delineation using wtdelineator
ecgDelin = delineateMultiLeadECG(np_data_12lead, 2000)
# Gather the [P-wave onset and end] of frontal leads in list
#   'frontalLeads'
frontalLeads = []
for idx in range(5):
    frontalLeads.append([[Pon, Pend] for
        Pon, P, Pend, QRSon, R, QRSend, Ton, T, Tend
            in ecgDelin[idx]])
# list of P-Waves found in each lead
n_pwaves = [len(lead) for lead in frontalLeads]
# Finding amount of P-waves found, correlating to most leads.
#   Looking to get as many usable leads as possible as result
sum = 0
max_n = 0
for length in n_pwaves:
```

```
    if sum < n_pwaves.count(length):
        sum = n_pwaves.count(length)
        max_n = length
frontalLeads = [lead for lead in frontalLeads
                    if len(lead) == max_n]
```

P-wave Voltage

```
leadI = np_data[:,0]
p_voltage = [max(leadI[Pon:Pend])-min(leadI[Pon:Pend]) for
               Pon, Pend in p_wave_ix]
```

P-wave Area

```
maxVoltage = [max(np_data[:,1][Pon:Pend]) for Pon, Pend
                    in p_wave_ix]
minVoltage = [min(np_data[:,1][Pon:Pend]) for Pon, Pend
                    in p_wave_ix]
p_area = [0.5*duration*voltage for duration, voltage in
            zip(p_duration, maxVoltage-minVoltage)]
```