Computers in Biology and Medicine 143 (2022) 105270

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Classification of *De novo* post-operative and persistent atrial fibrillation using multi-channel ECG recordings

Hanie Moghaddasi [a,*], Richard C. Hendriks [a], Alle-Jan van der Veen [a], Natasja M.S. de Groot [a,b], Borbála Hunyadi [a]

[a] Circuits and Systems, Delft University of Technology, Delft, the Netherlands
[b] Department of Cardiology, Erasmus University Medical Center, Rotterdam, the Netherlands

ARTICLE INFO

Keywords:
Atrial fibrillation
Poincaré
Vectorcardiogram
Atrial activity
Dominant frequency

ABSTRACT

Atrial fibrillation (AF) is the most sustained arrhythmia in the heart and also the most common complication developed after cardiac surgery. Due to its progressive nature, timely detection of AF is important. Currently, physicians use a surface electrocardiogram (ECG) for AF diagnosis. However, when the patient develops AF, its various development stages are not distinguishable for cardiologists based on visual inspection of the surface ECG signals. Therefore, severity detection of AF could start from differentiating between short-lasting AF and long-lasting AF. Here, *de novo* post-operative AF (POAF) is a good model for short-lasting AF while long-lasting AF can be represented by persistent AF. Therefore, we address in this paper a binary severity detection of AF for two specific types of AF. We focus on the differentiation of these two types as *de novo* POAF is the first time that a patient develops AF. Hence, comparing its development to a more severe stage of AF (e.g., persistent AF) could be beneficial in unveiling the electrical changes in the atrium. To the best of our knowledge, this is the first paper that aims to differentiate these different AF stages. We propose a method that consists of three sets of discriminative features based on fundamentally different aspects of the multi-channel ECG data, namely based on the analysis of RR intervals, a greyscale image representation of the vectorcardiogram, and the frequency domain representation of the ECG. Due to the nature of AF, these features are able to capture both morphological and rhythmic changes in the ECGs. Our classification system consists of a random forest classifier, after a feature selection stage using the ReliefF method. The detection efficiency is tested on 151 patients using 5-fold cross-validation. We achieved 89.07% accuracy in the classification of *de novo* POAF and persistent AF. The results show that the features are discriminative to reveal the severity of AF. Moreover, inspection of the most important features sheds light on the different characteristics of *de novo* post-operative and persistent AF.

1. Introduction

Atrial fibrillation (AF) is the most common and sustained arrhythmia in the heart, affecting 2% of the world's population, increasing the mortality rate and cost of health care [1]. One out of four individuals is expected to experience AF in their lifetime [1]. It is known that early detection of AF increases the chance of a timely treatment and improves life expectancy. AF is described by uncoordinated atrial activity that is represented on the electrocardiogram (ECG) by irregular RR intervals. Instead of having a single P wave in the early systolic time, AF is characterized by presence of fibrillatory waves or even absence of the P wave [2]. Cardiology guidelines classify AF based on the duration of AF episodes [3]. According to this definition, in paroxysmal AF, an AF episode lasts between 30 s and 7 days and returns to normal sinus rhythm (NSR) by itself, while in persistent AF, an AF episode lasts more than 7 days and it is not self-terminating [3–5]. Interestingly, atrial fibrillation is the most common complication after cardiac surgery in patients who never experienced AF [6]. These patients are commonly referred to as *de novo* post-operative AF (POAF) patients. The specific mechanism of *de novo* POAF is not yet fully understood. Hypotheses state that the patients develop AF after cardiac surgery due to reasons such as pericardial inflammation, atrial stretch and disharmony in the autonomic nervous system, to name a few [7]. All these factors shorten the refractory period of the atrial cells, making the wavelets propagate to the atrioventricular (AV) node at a higher rate, resulting in fast and irregular ventricular contraction [6].

* Corresponding author.
E-mail address: H.moghaddasi@tudelft.nl (H. Moghaddasi).

https://doi.org/10.1016/j.compbiomed.2022.105270
Received 12 October 2021; Received in revised form 24 January 2022; Accepted 24 January 2022
Available online 1 February 2022

Currently, the gold standard to diagnose AF is considered to be the interpretation of the 12-lead ECG by a trained physician [8]. However, based on visual inspection of ECG signals with AF episodes only, it is impossible for cardiologists to distinguish between *de novo* POAF and persistent AF. Specifically, if the patient suffers from AF for a longer time, there are more structural alterations of the atrial tissue (arrhythmogenic substrate). Furthermore, the stage of AF (severity of the arrhythmogenic substrate) when the patient develops AF cannot be determined by visual inspection of the ECG by physicians. In other words, AF episodes, do not show a visible difference between *de novo* POAF and persistent AF. This shows the need for an algorithm which can help physicians to detect the severity of AF.

The simplest level in the severity detection of AF is to classify ECGs into NSR episodes and AF episodes. Scientists addressed this issue by computer-aided diagnosis (CAD) systems. Generally, the methods can be classified into two main groups as handcrafted- and deep learning-based methods. The handcrafted features are extracted based on the prior knowledge which are fed to the classifiers. Due to the nature of AF, the handcrafted features can be classified into 4 groups; time-, frequency-, time-frequency and nonlinear-based methods. The time-based features mostly focused on the irregularity in the RR intervals [9] or the discrepancy in atrial activity [10]. The irregularity in the RR intervals is demonstrated by the features which rely on the heart rate variability (HRV) using Poincaré plots [11,12], density histogram of delta RR intervals [13], entropy measures [14,15], probability density estimation of the RR interval distribution [16] and calculation of statistics [17–20]. On the other hand, the discrepancy in the atrial activity is evaluated by the P-R interval variability, the P wave morphology similarity measure, the R-R Markov score [21] and absence of a P wave or presence of fibrillatory waves using piecewise linear functions [22].

But, over and above this, frequency analysis of the fibrillatory waves is quite informative. The fibrillatory wave ratio using a power spectral density [23] and fibrillatory wave spectrum [24] are common extracted features in this category. Although, time/frequency-based features have the capability of extracting irregularity of RR intervals and fibrillatory waves, techniques based on wavelet transform solve the low frequency/time resolution problem of the previous methods by jointly localizing in time and frequency. Atrial activity were correctly detected by features extracted from the correlation matrix of the wavelet coefficients [25] and discrete wavelet transform [26]. Furthermore, dynamic characteristics of AF is extracted by features as higher-order spectra (HOS) [27], Hurst exponent, largest Lyapunov exponent (LLE), and fractal dimension [28], quadratic sample entropy and LLE [29]. In contrast to the handcrafted features, most of the deep learning-based methods are independent from the discriminative features [30] or less-dependent on the handcrafted features [31]. At the same time, they provide less insight into the underlying electropathology of AF.

Although there exists extensive research for the classification of NSR episodes and AF episodes, this is not directly useful for diagnosing the severity of AF. This is due to the fact that both *de novo* POAF and persistent AF patients can have both NSR and AF episodes. In this paper, we tackle the challenging problem of diagnosing the severity of AF based on AF episodes. There is no clear, visually perceivable difference between an AF episode in a *de novo* POAF patient and an AF episode in a persistent AF patient. Our motivation to address this problem is summarized as follows. First, since *de novo* POAF is the first time that a patient develops AF, comparing its evolution to persistent AF might lead to insights into the exact mechanism of AF and its underlying electropathology. Moreover, being able to differentiate between short-lasting and long-lasting AF could help physicians to learn how to determine the stage of its development. In fact, knowing this is critical for AF treatment. Considering that AF is a progressive disease that can lead to stroke or heart failure, detection and treatment of AF in the early stages will decrease the mortality rate and healthcare costs. In addition, the development stage of AF is crucial for predicting the outcome of the treatment. An example of this is catheter ablation, the standard therapy for AF patients. In the case of paroxysmal patients, the success rate of ablation therapy is between 70% and 80% while for patients with persistent AF, this rate drops to 45–60% [32]. This shows why early detection of AF is of vital importance in the treatment process. Finally, this method provides a practical tool to find the severity of AF based on the AF episodes. This is valuable since the stages of AF can be detected at the time of the developing AF.

In this paper, we analyze long-term multi-channel surface ECG recordings. Our hypothesis is that there is a clear difference between *de novo* POAF and persistent AF. We test this hypothesis by proposing features that can differentiate between these two groups. Analyzing these differences may give us insight into the physiological changes that underlie the occurrence of AF and, subsequently, reveal the underlying differences.

The core of our method therefore consists of careful feature engineering preceding machine learning-based classification. To capture morphological and rhythmic differences, multi-channel ECGs can be analyzed in various ways, e.g., the time domain, the spatial domain, and the frequency domain. We propose three corresponding groups of features. In the time domain, we propose rhythm-based features where we look at the irregularity of the RR intervals, an indirect indicator of the atrial activity. The second set of features is based on the vectorcardiogram. Here we extract joint spatial/temporal aspects of the heart's dynamics in the cardiac cycles. This is implemented based on image descriptors on the 3D structure of the electrical activity of the heart. Finally, in the third group of features, frequency-domain aspects of the atrial activity during AF episodes are analyzed. The main novelty of the proposed method is hidden in the feature extraction. Firstly, we have introduced two novel features based on the different usage of autoregressive modeling and dominant frequency. Secondly, to the best of our knowledge, this is the first paper that detects the severity of AF by analyzing vectorcardiogram images using image processing techniques which unveils the new aspects of this disease. Furthermore, from the clinical point of view, we have introduced the first method of classification between *de novo* POAF and persistent AF which had not been done before.

The rest of this paper is organized as follows. In Section 2, we introduce our method including feature extraction, feature reduction and classification. The performance of our proposed method is demonstrated in Section 3. In Section 4, we analyze the feature values extracted from *de novo* POAF and persistent AF patients, to gain insight into the most essential differences between their underlying physiology. Finally, we draw the conclusions in Section 5.

2. Methodology

We hypothesize that differentiating *de novo* POAF and persistent AF can be done by finding features that capture the morphological and rhythmic changes in the ECG signal. With multi-channel ECG data, beat-to-beat variations in morphology are visible in patterns in the vector cardiogram. Rhythmic changes follow from surveying variations in the RR intervals (in time domain), and from analyzing the dominant frequencies corresponding to atrial activity.

In Fig. 1 we show a high-level block diagram of our approach outlining the different steps required for classification. The core of this work is the development of features discriminating between *de novo* POAF and persistent AF, as these features might provide more insight in how the development of AF can be monitored and how the development stage of AF can be determined in more detail. However, as our data originates from continuous recordings, noise and artifacts will be present. Prior to discussing the actual features, we will therefore discuss the data acquisition and the required pre-processing in Sections 2.1 and 2.2, respectively. As we hypothesize that differentiation between *de novo* AF and persistent AF lies in morphological and rhythmic beat-to-beat changes, we will develop in Sections 2.3–2.5 three groups of features capturing these aspects. Subsequently, we discuss the feature selection

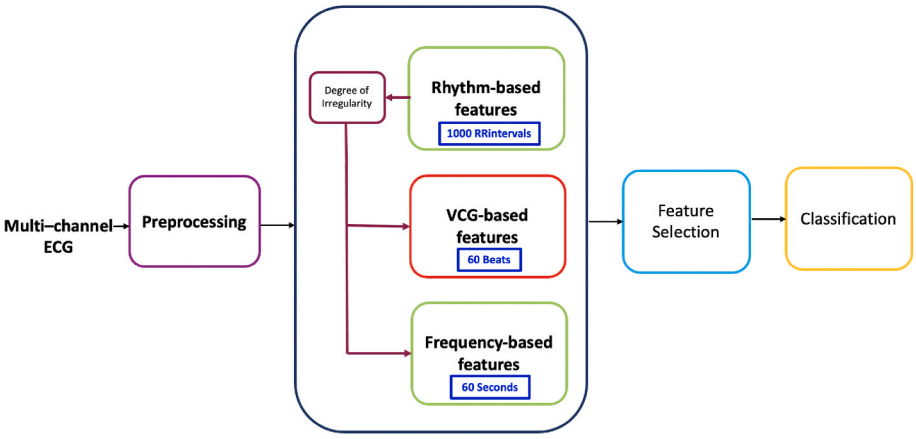

Fig. 1. Block diagram of the proposed method.

stage and classification in Sections 2.6 and 2.7, respectively.

2.1. Data acquisition

We enrolled in total 151 patients of which 99 were categorised as *de novo* and 52 were categorised as persistent AF patients. The telemetry data was collected at the Erasmus Medical Center (EMC). The data consists of 12-lead ECG signals with recording durations between 72 h and 120 h, at a sample rate of 200 Hz. All patients were labeled as *de novo* POAF or persistent AF by physicians at the EMC.

2.2. Pre-processing

The process of recording the ECG signals in the telemetry data takes several days per patient. During this period, the ECG leads are connected to the patient's body and it is likely that the leads sometimes get disconnected during the recording. Moreover, breathing artifacts, electromyography interference and power-line interference are other inevitable noise sources. To remove the effect of these perturbations, we filter the data with a Butterworth band-pass filter in the frequency range between 0.33 Hz and 30 Hz [33–35]. In addition, due to the high-amplitude noise, the quality in some segments of the long-term recordings is poor. These segments are therefore excluded from the signal. To do so, we divide the ECG signals in each lead in the pre-processing step into 60s segments and apply the band-pass filter to all segments. For each segment and channel, we estimate the signal-to-noise ratio (SNR) by calculating the power of the denoised signal divided by the power of the residual noise as

$$\mathrm{SNR} = 10\log_{10}\frac{\sum_{n=0}^{N-1} s_d^2(n)}{\sum_{n=0}^{N-1}\left(s(n)-s_d(n)\right)^2} \tag{1}$$

where N is the number of samples in a segment, and $s(n)$ and $s_d(n)$ are the original and the denoised signals, respectively. After that, we remove the segments where the ECG has an SNR < 10 dB. Notice that the SNR is determined based on the noise outside the frequency range of 0.33–30 Hz. Implicitly we thus assume that the noise outside this frequency range is also representative for the quality of the ECG signal inside this frequency range. The retained segments are normalized with respect to the maximum absolute value of the signal, to have an amplitude between -1 and $+1$. For the features which need R peaks detection, we have done another step in the pre-processing. By calculating SNR as in Eq. (1) for all ECG channels, we select the ECG channel with the highest SNR for performing R peak detection. Finally, a wavelet-based ECG delineator algorithm [36] is applied to the filtered ECG signal to detect the R peak in each cardiac cycle in both forward and reverse directions.

2.3. Rhythm-based features

One way in which AF affects the ECG is by increasing the irregularity of the RR intervals. Generally these intervals become shorter and less predictable. Fig. 2A shows an example of the RR intervals of a *de novo* POAF, and Fig. 2B shows an example of the RR intervals of a persistent AF. For this figure, for the *de novo* POAF patient a data segment was selected where AF is present. It is clearly seen that regular (and slower) beats are interlaced with irregular and faster beats. For persistent AF, the regular beats are absent.

Another way to visualize this is through Poincaré plots [11]. This is a recurrent scatter plot that allows to judge the correlation structure present in a time series, in this case a sequence of N RR intervals with the individual lengths given by $I_1, I_2, \ldots, I_N$. The plot shows the points $[I_k, I_{k+1}]$, for $k = 1, \ldots, N-1$. As example, Fig. 3A shows a Poincaré plot of a normal sinus rhythm (NSR) episode of a *de novo* POAF patient, Fig. 3B shows a Poincaré plot of an AF episode a *de novo* POAF patient, and Fig. 3C shows a Poincaré plot of a patient in persistent AF. Clearly, in the latter case the Poincaré plot does not represent any specific pattern and the points have an irregular distribution, while during *de novo* POAF, two different patterns are observed. In the area indicated by the symbol O, points are concentrated on an ellipsoid oriented around a diagonal line (for which $I_k = I_{k+1}$), while for the remaining points in the Poincaré plot no regular pattern is observed. The latter cloud of points is absent for the NSR episodes.

In the related literature, several parameters have been proposed to describe the structure in such plots [37], often refer to as SDNN (standard deviation of NN intervals e.g., the standard deviation of I_k, with outliers removed), SD1 and SD2 (the axis lengths of a fitted ellisoid), SDSD (the standard deviation of successive differences $I_{k+1} - I_k$), and RMSSD (the root mean square of successive differences), etc. As discussed in Ref. [38], these are all related, and we offer the following perspective from statistical system theory to explain this further.

Consider the measured RR intervals $I_1, \ldots, I_N$ as a realization of a wide sense stationary random process. Let $\bar{I}$ denotes the sample mean, that is,

$$\bar{I} = \frac{1}{N}\sum_{k=1}^{N} I_k. \tag{2}$$

We will work from now on with the zero mean sequence $x_k := I_k - \bar{I}$.

The Poincaré plot is related to the modeling of x_k by a first-order autoregressive (AR) model, AR(1), as

$$x_{k+1} = a\,x_k + e_k, \tag{3}$$

where e_k is a zero mean Gaussian random process with variance σ^2, and a is the AR coefficient ($|a| \leq 1$). If $a = 1$, $\sigma = 0$, then $x_{k+1} = x_k$ and the RR intervals are constant and perfectly predictable. If $a = 0$, then the RR

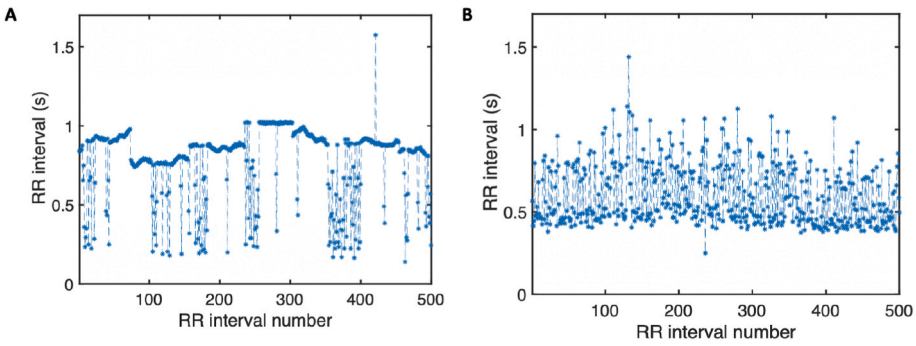

Fig. 2. RR intervals for A) *de novo* POAF, B) persistent AF.

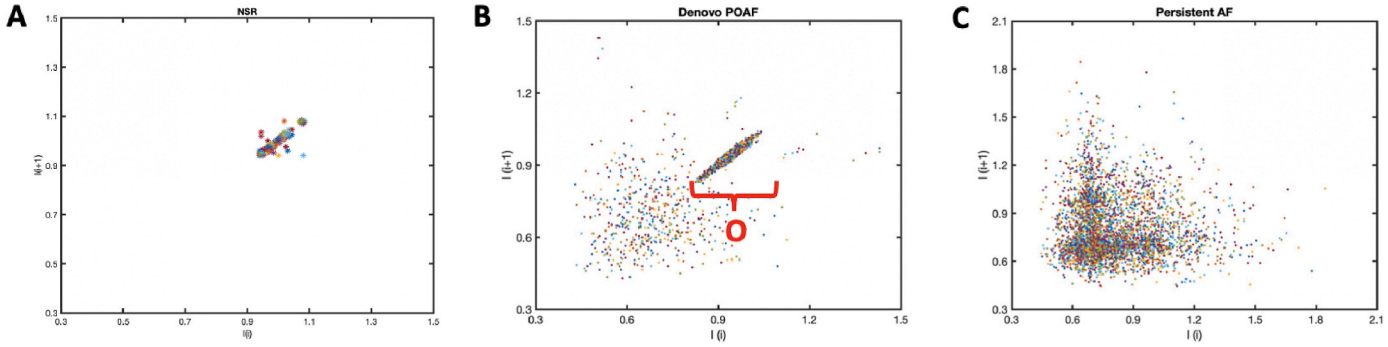

Fig. 3. Poincaré plots: A) NSR intervals in *de novo* POAF, B) irregular intervals in *de novo* POAF, C) persistent AF.

intervals are completely uncorrelated.

For some upper bound L, not too large, we form the sample auto-covariance sequence

$$\widehat{r}_\ell = \frac{1}{N-L}\sum_{k=1}^{N-L} x_{k+\ell}x_k, \qquad \ell = 0, 1, \ldots, L. \tag{4}$$

These are estimates of $r_\ell = E[x_{k+\ell}x_k]$. Then the AR(1) model satisfies [39].

$$\begin{cases} r_0 &= \dfrac{\sigma^2}{1-a^2} \\ r_1 &= a\, r_0 \end{cases} \Leftrightarrow \begin{cases} a &= \dfrac{r_1}{r_0} \\ \sigma^2 &= \dfrac{r_0^2 - r_1^2}{r_0^2} \end{cases}$$

We can also construct the 2×2 correlation matrix

$$\boldsymbol{R} = \begin{bmatrix} r_0 & r_1 \\ r_1 & r_0 \end{bmatrix} = \frac{\sigma^2}{1-a^2}\begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}.$$

The eigenvalue decomposition of $\boldsymbol{R}$ can be computed in closed form as

$$\boldsymbol{R} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_d^2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

with

$$\sigma_a^2 = r_0 + r_1 = \frac{\sigma^2}{1-a}$$
$$\sigma_d^2 = r_0 - r_1 = \frac{\sigma^2}{1+a}.$$

This latter parametrization is directly visible in the Poincaré plot: for an AR(1) model, the points are scattered on an ellipsoid with center $[\bar{I},\bar{I}]$, with the long axis in the direction $[1,1]$ with length σ_a, and with the short axis in the orthogonal direction $[1, -1]$ with length σ_d. Further, σ_a is equal to SD1, while σ_d corresponds to SD2, SDSD, and RMSSD.

Thus, we have 3 equivalent parametrizations of the AR(1) model: $(\bar{I}, r_0, r_1)$, $(\bar{I}, a, \sigma^2)$, and $(\bar{I}, \sigma_a, \sigma_d)$.

To measure the irregularity, an index δ is defined as

$$\delta = \frac{\sigma_d^2}{\bar{I}} \tag{5}$$

which is the difference of the time-averaged correlation values with zero and one time lags normalized by the mean of the RR intervals. For regular RR intervals (e.g., NSR episodes), it is expected that all the points are located close to a central point which means that δ is almost zero. As the irregularity increases, δ increases. We use this index to filter out the NSR episodes in the classification.

At this point, a relevant question is also whether AR(1) is a good model choice at all. We could study the fit to higher-order models, AR(ℓ), for $\ell = 1, \ldots, L$, and related performance metrics such as final prediction error (FPE) [40]. Instead, we will consider more general auto-regressive moving average (ARMA) models. For some p, larger than the model order we wish to select, construct the $p \times p$ Hankel matrix

$$\boldsymbol{H} = \begin{bmatrix} r_1 & r_2 & \cdots & r_p \\ r_2 & r_3 & \cdots & r_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ r_p & r_{p+1} & \cdots & r_{2p-1} \end{bmatrix}.$$

Then system theory tells us that for an ARMA(ℓ) model, the rank of $\boldsymbol{H}$ is ℓ. For example, for an AR(1) model,

$$\boldsymbol{H} = r_0\begin{bmatrix} a & a^2 & a^3 & \cdots \\ a^2 & a^3 & a^4 & \cdots \\ a^3 & a^4 & a^5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

which is clearly of rank 1. To detect the rank, consider the singular values of $\boldsymbol{H}$ sorted in non-increasing order, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. If the rank of $\boldsymbol{R}$ is ℓ, then $\lambda_\ell > 0$ while $\lambda_{\ell+1} = 0$.

As feature to detect rank 1 in our data sets, we propose to use $p = 5$,

and define the feature

$$\varrho = \frac{\lambda_1}{\lambda_2}. \tag{6}$$

If a first order model is a good fit, then $\varrho \gg 1$. Fig. 4 shows as example plots of the eigenvalues of $\boldsymbol{H}$ for *de novo* POAF and persistent AF, respectively. It is seen that for *de novo* POAF $\lambda_1 \gg \lambda_2$, while for persistent AF λ_1 is not much larger than λ_2. Thus, for *de novo* POAF, ϱ is much larger than for persistent AF, and a rank 1 model seems appropriate. Comparing to Fig. 3B, we expect that points in the narrow ellipsoidal area indicated by the symbol O dominate the model structure.

Altogether, to compute the rhythm-based features, we first exclude the NSR episodes by excluding segments for which δ is almost zero (corresponding to a highly regular/predictable RR interval sequence). By analyzing our dataset, a threshold $\delta > 0.01$ is defined to exclude NSR episodes. For the classification, considering the shortest AF episode in our dataset, we used a window that contains 1000 RR intervals (from the most irregular part of the signal as measured by the δ parameter) and extracted four features from the best quality lead: $\bar{I}$, σ_a, σ_d, and ϱ.

2.4. Vector cardiogram-based features

In the previous section, we extracted features from the time domain. In this section, we investigate spatial information of the ECG signals.

Different degrees of AF are expected to represent themselves in terms of variations across time on the parameters required to model the different ECG components. This becomes even more prominent if we take multiple leads into account, as morphology may vary across different leads. In order to summarize multi-channel ECG information in a compact way, we make use of the vector cardiogram (VCG). The VCG is an alternative representation of the multi-channel ECG information that allows to track the electrical activity of the heart along the three orthogonal body planes: left-right (x), head-to-feet (y) and front-back (z). Using Frank's transformation [41,42], the 3-lead VCG ($\boldsymbol{v} = [V_x, V_y, V_z]^T$) is obtained using 8 leads out of the 12-lead ECG by

$$\boldsymbol{v} = \boldsymbol{F}\,\boldsymbol{e} \tag{7}$$

where $\boldsymbol{e} = [V_1, V_2, V_3, V_4, V_5, V_6, \mathrm{I}, \mathrm{II}]^T$ and $\boldsymbol{F}$ is a 3×8 transformation matrix derived from Frank's equation [41].

As an example, Fig. 5A and Fig. 5B show the orthogonal leads as a function of time for an NSR episode and an AF episode in a persistent AF patient, respectively. The corresponding 3D VCGs are shown in Fig. 5C and D. For NSR, each cardiac cycle creates a complete P loop (red), QRS loop (green) and T loop (blue). However, for persistent AF, the P loops are not clearly visible, but replaced by a chaotic trajectory before the start of the QRS loops. Moreover, for persistent AF, beat-to-beat variations of the QRS loop are greater than for NSR. The most important differences between the 3D VCG of NSR and persistent AF are the sub-patterns within the image representing the electrical activity of the heart along the three orthogonal body planes. A suitable discriminator is therefore a feature describing the local patterns in this 3D VCG image.

To capture these differences, we propose to use local binary pattern (LBP) analysis [43] on a segmented version of the 3D VCG image. LBP is an image descriptor that has been successfully used to describe texture in various applications ranging from face detection and detection of facial expressions [44,45], to diagnosing heart diseases [46,47].

The main idea behind LBP analysis is to first capture and categorize the local texture of the image around a given pixel using the so-called LBP operator, and subsequently summarize the information from all pixels using a histogram. The LBP operator computes an 8-bit binary code word for each pixel by comparing its grayscale value to its 8 neighbors in a 3×3 neighborhood (see Fig. 6). If the intensity of the given pixel is smaller than that of its ith neighbor, the corresponding bit is set to '1', and otherwise it is '0'. In the original LBP algorithm [43], this leads to a total of 2^8 possible binary patterns. Then, the number of occurrences of each pattern is counted to build a histogram. This histogram describes the distribution of all possible local patterns in the image. However, not all 256 possible patterns are equally interesting. Firstly, many of these patterns can be obtained from each other by a circular shift. There are only 36 unique rotationally invariant patterns (see Ref. [48] for the visualization of all these patterns). Secondly, it has been shown empirically that a vast majority of all patterns found in real images share a certain property, namely, that they are 'uniform'. Here, uniformity means that a circular pattern has at most 2 transitions from black to white and vice versa. For example, the pattern 00 000 001 (2 transitions, considering a cyclic extension where the last bit is also compared to the first) is uniform while the pattern 11 001 001 (4 transitions) is non-uniform. Interestingly, the most frequent uniform patterns turn out to represent important microfeatures in the image such as dots or edges. Out of the 36 rotationally invariant patterns only 9 are uniform. These patterns are shown in Fig. 7. The first and last pattern capture bright spots and black spots or homogeneous surfaces, respectively. The other seven patterns capture edges between a bright and a dark surface. Finally, in our adaptation of the uniform rotation invariant LBP (URILBP) we construct a histogram by counting the number of occurrences in these categories. As such, our VCG-based URILBP feature has a length of 9.

The original 3D VCG image is a large binary volume. Before applying URILBP, we first discretize and segment this volume in such a way that the resulting set of grayscale images individually capture meaningful parts of the VCG. We first discretize the 3D image into 2D planes along the x, y, and z-direction with a resolution of 100 pixels per unit of normalized ECG voltage. As a result, binary images (2D planes) are

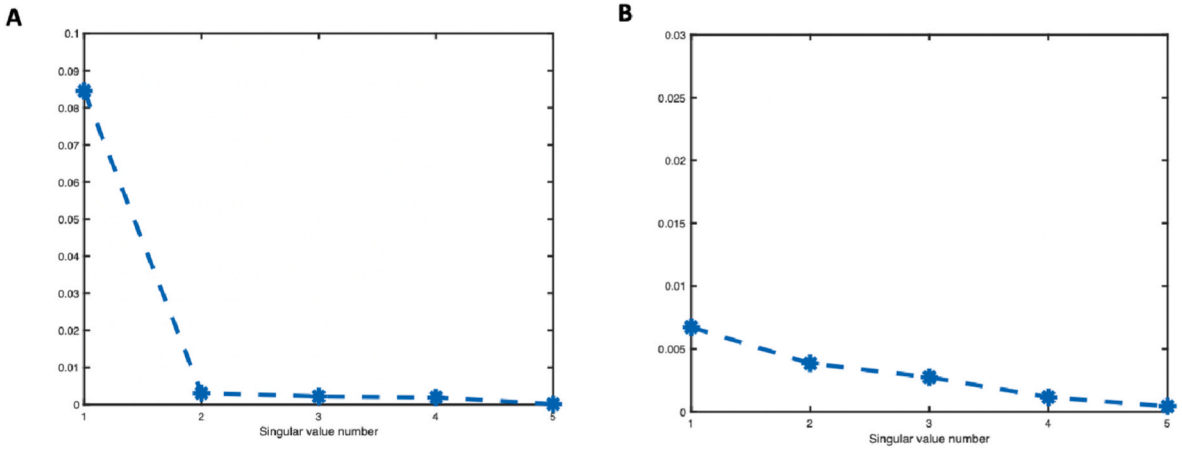

Fig. 4. Singular values of matrix $\boldsymbol{H}$: A) *de novo* POAF, B) persistent AF.

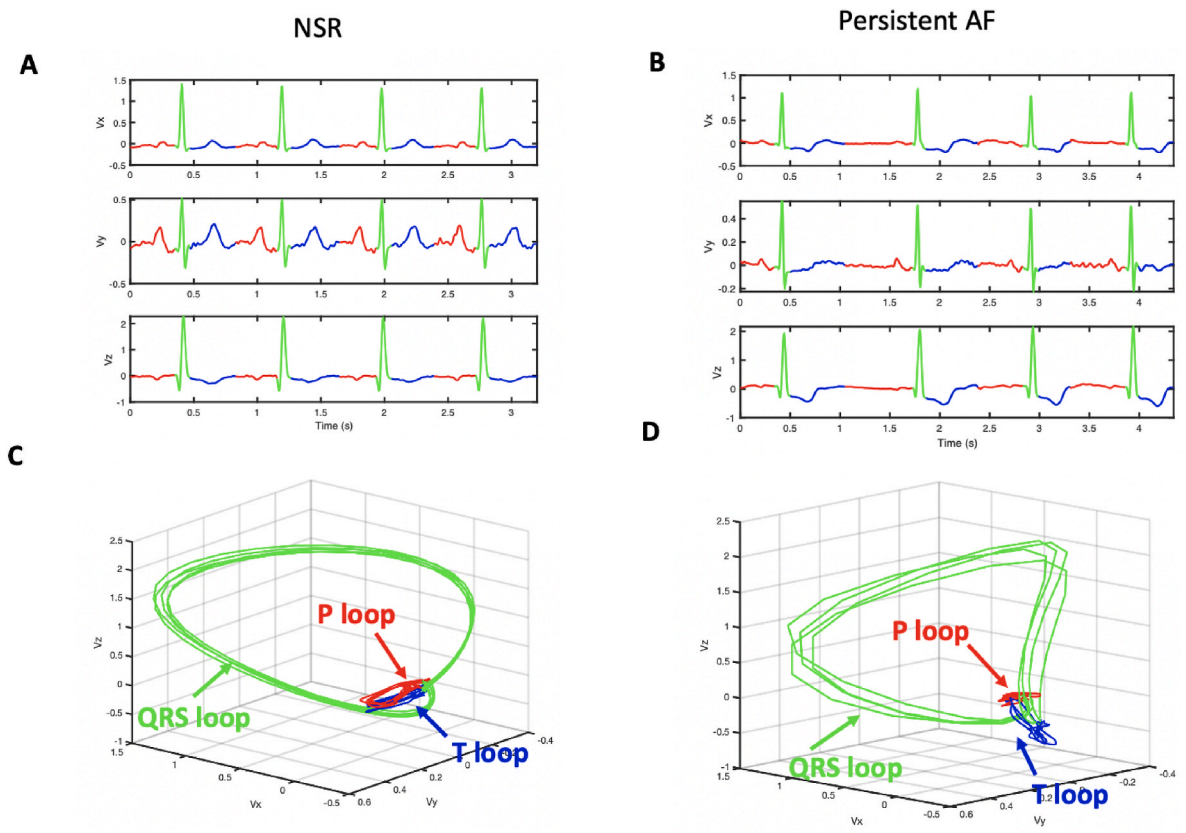

Fig. 5. A) Orthogonal leads for an NSR episode, and B) for a persistent AF patient; C) corresponding 3D VCG for the an NSR episode, and D) for the persistent AF patient.

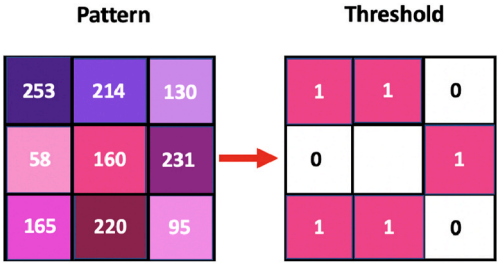

Fig. 6. LBP operator.

created along all three directions. Next, the binary images are divided into 5 batches in each direction. The images in the batches are summed together, resulting in 5 2D grayscale images in each direction, i.e. 15 planes in total. We emphasize that the major motivation behind this preprocessing is to segment the 3D VCG into P loops, QRS loops and T loops. In Figs. 10 and 5 planes in x-direction are shown. Plane 3 and plane 4 mainly contain the P loops and T loops, respectively, while QRS loops can be detected in plane 1, plane 2 and planes 5. As the P and T loops occupy just part of the plane, the planes are further divided into 9 non-overlapping sub-images and the URILBP histogram is calculated for each sub-image. When the loops follow the same or similar trajectory across subsequent heartbeats, the subimages (in the direction perpendicular to the local trajectory) will contain a majority of black pixels with bright areas (see for example bottom left of plane 3). The edges of such areas may be captured by bins 4–8 in the URILBP histogram. Conversely, chaotic trajectories will result in numerous discrete bright dots (see for example bottom right of plane 1), that can be captured by bins 2–3 of the URILBP histogram.

The nine histograms extracted from the sub-images are concatenated

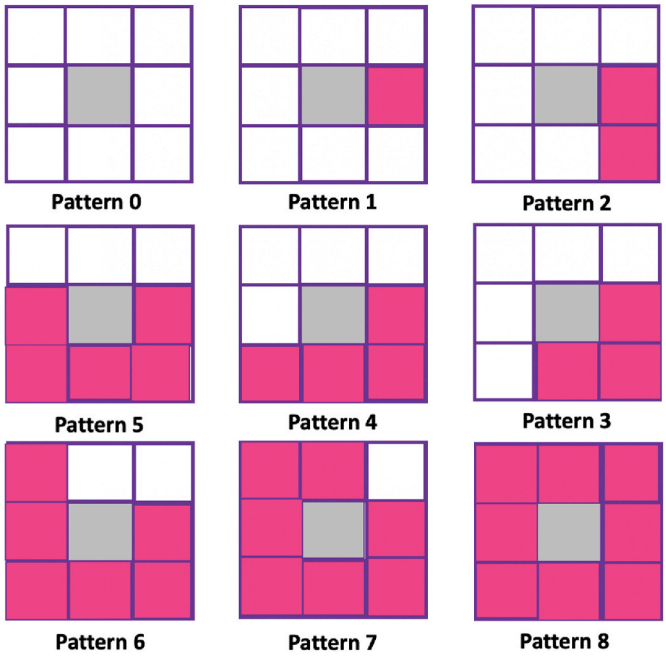

Fig. 7. URILBP: There are in total 9 possible uniform rotationally invariant local binary patterns.

to create a feature vector for that plane (see Fig. 8). Finally, the feature vectors of the 15 planes are concatenated into one final feature vector (see Fig. 9). We refer to the resulting procedure as 3DLBP.

To compare the 3DLBP between *de novo* AF and persistent AF, we

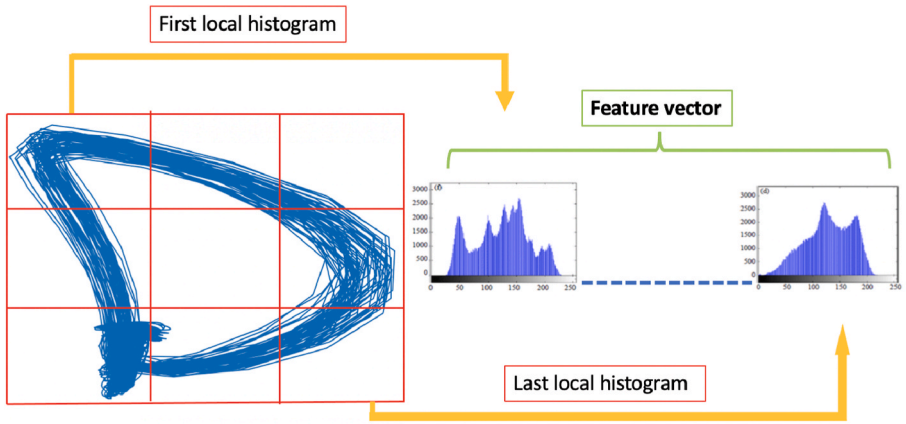

Fig. 8. The URILBP feature vector of an image consists of a concatenation of the histogram of URILBP scores of each sub-image.

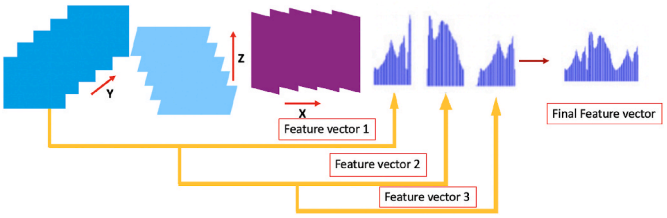

Fig. 9. 3DLBP feature vector.

divide the filtered ECG signals into frames of 60 beats. We select the frame with the highest degree of irregularity, as measured by the δ parameter (see Eq. (5)). This frame is selected for the following reasons. First, the highest δ frame is representative of the most irregular parts of the signal, so by using 3DLBP on this frame, we can extract the texture information of the VCG trajectory in the most irregular segment of the signal. Therefore, this enables us to compare the VCG trajectory in the AF episodes in *de novo* AF and persistent AF. Moreover, this will suppress the repetitive information of NSR episodes in *de novo* patients and AF episodes with the same trajectory in the VCG. In other words, this enables us to compare a small portion of the signals that is the most characteristic of the ECG in AF patients.

2.5. Frequency-based features

The time domain and the spatial domain information are extracted in Section 2.3 and Section 2.4. As a third feature class, we propose to use specific frequency components of *de novo* POAF and persistent AF. During AF episodes, multiple wavefronts propagate simultaneously through the atrium. As a result, a surface ECG will show a broader frequency spectrum. Indeed, in NSR episodes the frequency range of atrial activity is typically between 0.6 Hz and 1.5 Hz, while during an AF episode the atrial activity is typically in the frequency range between 4 and 9 Hz [49,50]. We hypothesize that features which are related to the frequency components of atrial activity might also provide a distinction between *de novo* POAF and persistent AF.

To analyze the frequency content in the range 4–9 Hz, we follow the approach presented in Ref. [51]. Namely, we use a stationary wavelet transform with a Daubechies-5 mother wavelet to obtain the power spectrum $S_l^D(f)$ of the detail coefficients d_l at the lth level. In order to cover the desired frequency range, we first upsample the signal to 500 Hz and then use $l = 6$ levels of decomposition. The stationary wavelet transform has been used as a bandpass filter to extract the frequency range 4–9 Hz and then a Fourier transform is applied on the detail coefficients d_l at $l = 6$ to obtain the $S_l^D(f)$.

Then, we propose to find the dominant frequency (DF), which carries the fundamental frequency of the signal and can be calculated by finding the maximum peak in the power spectrum. In addition, the width of the power distribution profile of the signal, can be captured by the average power in the bandwidth of interest [51]. Therefore, we define the DF-to-average-power ratio, for the lth level, as

$$\rho = \frac{\arg\max_{f \in F} S_l^D(f)}{\int_{f \in F} S_l^D(f)\, df}. \tag{8}$$

To compare the DF-to-average-power ratio between *de novo* POAF and persistent AF, we divide the filtered ECG signals into 60 s segments and calculate the δ parameter according to Eq. (5). Then, we select the frame with highest δ (this indicates the most irregular and thus the most AF containing segment), because we want to compare the AF episodes in *de novo* POAF and persistent AF, while avoiding NSR episodes. This will increase the robustness of the classifier since frequency information in the AF episodes are only compared among each other. Since leads V_1, aVF and III have the largest atrial contribution in the ECG signal [52], we determine this feature from these three leads.

To compare the DF and DF-to-average-power ratio in *de novo* POAF and persistent AF, we divide the filtered ECG signals into 60 s segments and select the segments with the highest δ index. Then, we sub-divide

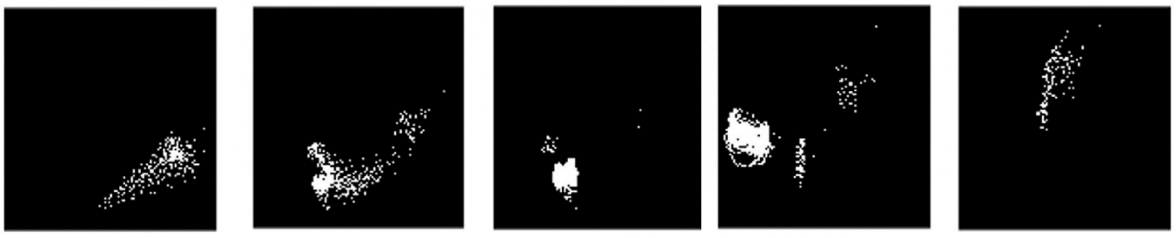

Fig. 10. Projected planes at x-direction.

each segment into 5 s non-overlapping frames. The estimated power spectral densities (PSDs) for all consecutive time frame of the detail coefficient in $l = 6$ are plotted on top of each other in gray for a *de novo* POAF patient and a persistent patient in Fig. 11A and Fig. 11B, respectively. The average PSD over time is shown in blue. Persistent AF has a larger DF with wider power distribution, while *de novo* POAF achieves a lower DF with narrower power distribution. Frequency components of atrial activity (AA) in persistent AF vary more than in *de novo* POAF. In the power spectrum for persistent AF, the peak is therefore lower and the width larger. In persistent AF, more frequency components with a lower signal amplitude are thus involved. Moreover, the frame-to-frame variation of the maximum peak of the PSD in *de novo* POAF is smaller than for persistent AF. This can be explained by the hypothesis that during persistent AF, more inconsistent wavelets travel across the atrium which is represented on the surface ECG by the multiple frequency components with most likely slightly different frequencies. Looking at the spectrograms (Fig. 11. C and Fig. 11. D), DF in *de novo* POAF is more consistent across time frames than persistent AF. It shows that by considering the whole sequence of spectra across 60 s, there are more variations in the activated frequency components in persistent AF.

2.6. Feature selection

When the number of features in a classification task is too large or/and features are correlated, feature selection is a critical pre-processing step for selecting the features with the highest relevance. This not only might reduce the computational burden of training the classifier but in many instances boosts its performance. In this work, we use ReliefF as our feature selection method [53]. ReliefF is a filter-based feature selection method which finds weights for the features. In the ReliefF algorithm, a feature from a random training sample is selected as R_i and then the algorithm searches for its two nearest neighbors as nearest hit H_j (from the same class) and nearest miss M_j (from the different class) [53]. The importance of a feature is represented by a weight. The weight increases if the feature and miss have different values while it decreases if the feature and hit have different values. This algorithm is repeated m times where m is a user-selected parameter. The highest weighted features are selected to train the classifier.

2.7. Classification

The performance of the extracted features is evaluated by implementing two supervised methods. In sections (2.7.1) and (2.7.2), we explain these methods used to classify patients into *de novo* POAF or persistent AF.

2.7.1. Support vector machine

With Support Vector Machines (SVMs), a decision boundary with a maximum margin is found for separating data points of different classes. Although an SVM is originally a linear classifier, the use of kernel functions enables SVM to perform nonlinear classification. Given a training set of P points as p_i: $i = 1, 2, \ldots, P$ with defined labels q_i, either -1 or $+1$, SVM classifies a test data p as [49].

$$f(p) = sgn\left(\sum_{i=1}^{P} \beta_i q_i K(p_i, p) + o\right) \tag{9}$$

where *sgn* is the sign function, β_i are Lagrange multipliers, K is the kernel function and o is the bias of the hyperplanes. In this work, we use the radial basis function (RBF) kernel function where we fixed the kernel parameter $\sigma = 1$. We empirically found that for our dataset this value represented a good trade-off between flexibility, i.e, higher testing accuracy, and the risk of overfitting.

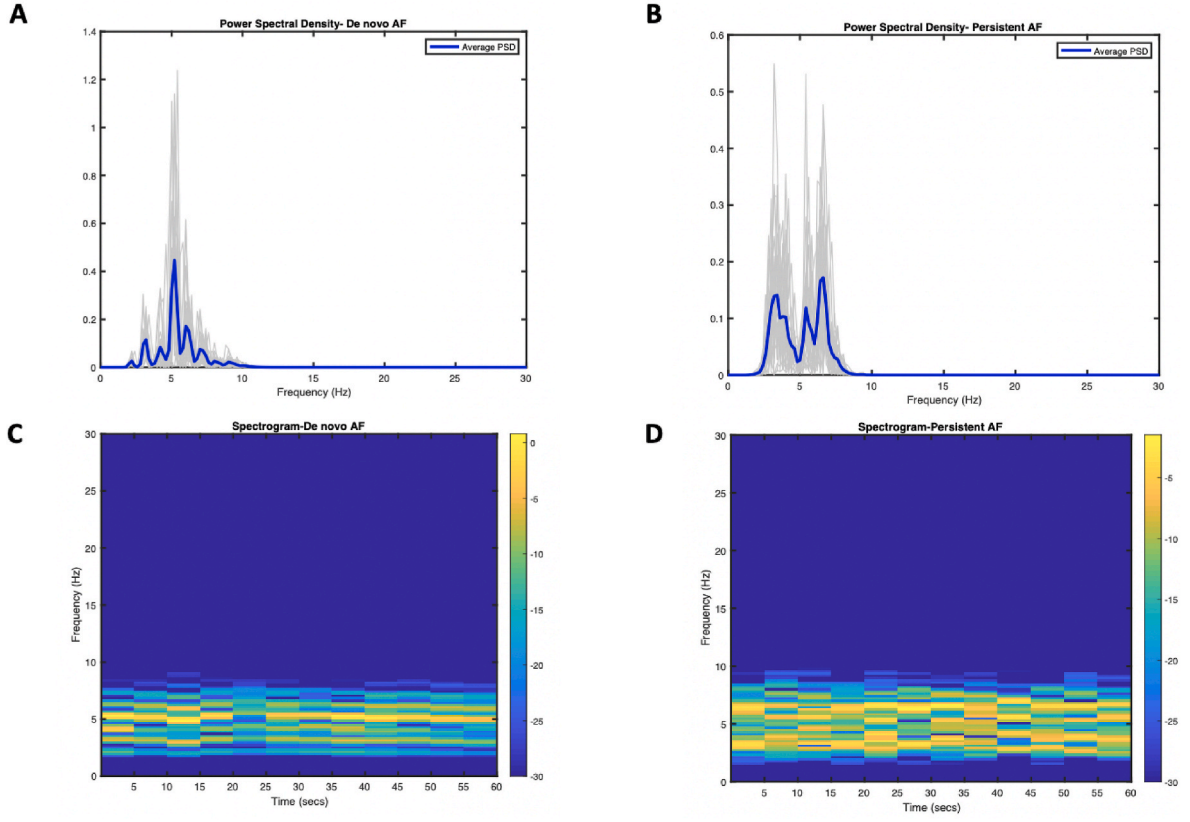

Fig. 11. Power spectral density and spectrogram, A) PSD of 60 s in a *de novo* POAF (gray) and the average PSD (blue), B) PSD of 60 s in a persistent AF (gray) and the average PSD (blue), C) spectrogram in a *de novo* POAF, D) spectrogram in a persistent AF.

2.7.2. Random forest

Random Forest (RF) is an ensemble learning method which is used for classification or regression problems. RF combines multiple decision trees for the final result in which each decision tree consists of three types of nodes, namely, the root node, inner nodes, and leaf nodes. Each decision predicts a result and the final result is produced by the majority voting among all trees. The advantage of using RF in our problem is that it solves the overfitting problem and the problem of unbalanced datasets and normalization of features is not necessary [54]. We trained a bagged ensemble of 150 classification trees with the maximum number of splits set to 13, using the highest weighted features.

3. Results

3.1. Feature analysis

The performance of the extracted features is demonstrated in this section. As in most *de novo* POAF patients AF develops after 48 h (day three), we extracted features starting from the third day (48 h–72 h) to observe AF episodes in patients. Considering the three groups of features (i.e., rhythm-, VCG- and frequency-based features) and after selecting the most irregular segments, we have used 121 979 s, in total. More specifically, we used 84 883 s from *de novo* POAF patients and 37 096 s from persistent AF patients. In the rhythm-based features, 1000 RR intervals have been used per patient which vary between 372s and 938s. For the VCG-based features, we used 60 beats per patient which vary between 24s and 57s. In the frequency-based features, we used 60s per patient. Four features, namely, $\bar{I}$, σ_a, σ_d, ϱ are extracted from the rhythm-based feature, in total. From the frequency-based features, ρ is calculated for the leads V_1, *aVF*, and III (three features). For the VCG-based features, using URILBP for the most irregular (highest δ) frame of the 3D VCG, results in a feature vector with length 1215. Therefore, the initial length of the feature vector, for each patient, is 1222. Negative weights estimated by ReliefF are not suitable predictors for the classifier [53], so by using ReliefF, we then selected the 32 highest weighted features (positive ones) for training the classifier, which turned out to be all rhythm-based features ($\bar{I}$, σ_a, σ_d, and ϱ), ρ from lead V_1, and some bins of the URILBP histogram.

Fig. 12 shows the box plots of some of the selected features. In Fig. 12A and Fig. 12B three rhythm-based features (σ_a, σ_d and $\bar{I}$) and two ratios (ϱ and ρ) are shown, respectively. Four selected bins of the VCG-based features are shown in Fig. 12C. Looking at Fig. 12A, in the *denovo* POAF patients, the σ_a, σ_d and $\bar{I}$ are larger than for persistent AF. Comparing ϱ and ρ (Fig. 12B) in *de novo* POAF and persistent AF, ϱ in *de novo* POAF is larger than for persistent AF which shows that the rank 1 model is appropriate for *de novo* POAF. Also, ρ is larger in *de novo* POAF than in persistent AF, as there are more frequency components in persistent AF than in *de novo* POAF. For the VCG-based features, 27 bins are selected by the ReliefF algorithm. Here, we have interpreted a few of these bins (four patterns which are shown in Fig. 12C) for the following reasons. At first, some of the VCG bins are discriminative independently and the rest are jointly insightful. In this section, we focus on the independent bins. Second, VCG-based features are extracted from three directions (X–Y planes, X-Z planes and Y-Z planes). So, examples from all three directions are shown. Third, these bins have the highest weights in the ReliefF algorithm. Therefore, we have interpreted the most discriminative bins. The selected features are mainly concentrated on the sub-images that correspond to the P loops area. In Fig. 12C, patterns 5 and 6 belong to the P loops area on the X–Y plane. The patterns represent the edges in the image. The edges are indicators for a particular pattern in an image. These patterns can show the difference between a loop and a chaotic pattern. In the *de novo* patients, the number of pixels that represent patterns 5 and 6 is larger than in persistent AF patients. It shows that there are more edges for *de novo* patients than for persistent AF. Conversely, patterns 1 and 2 represent single dot and double dots in the 3×3 neighborhood, respectively. In Fig. 12C, pattern 1 is a bin on the P loops area on the X-Z plane and pattern 2 is on the Y-Z plane. In persistent AF, the number of pixels that represent patterns 1 and 2 is larger than in *de novo* POAF patients. The more chaotic the pattern, the more dots there are on the LBP image. In other words, these box plots represent that in the bins that correspond to edges (i.e. regular loops), the number of pixels in the *de novo* POAF is higher than in persistent AF, while in bins that correspond to dots (i.e. chaotic patterns), the number of pixels in persistent AF is higher than in *de novo* POAF.

For the validation of the classifier, we report the mean (M) and standard deviation (SD) of 5-fold cross validation on 151 patients (99 *de novo* POAF patients and 52 persistent AF patients). To keep the same proportion for each group in the training set as in whole dataset, we applied 5-fold splitting on each class separately. In 5-fold cross validation testing, we have under-sampled the majority class (i.e., *de novo* POAF) to have the same number of patients to evaluate in both classes [55].

3.2. Performance evaluation parameters

To check the performance of the classifier, we evaluate our results using common performance metrics such as accuracy (ACC), sensitivity

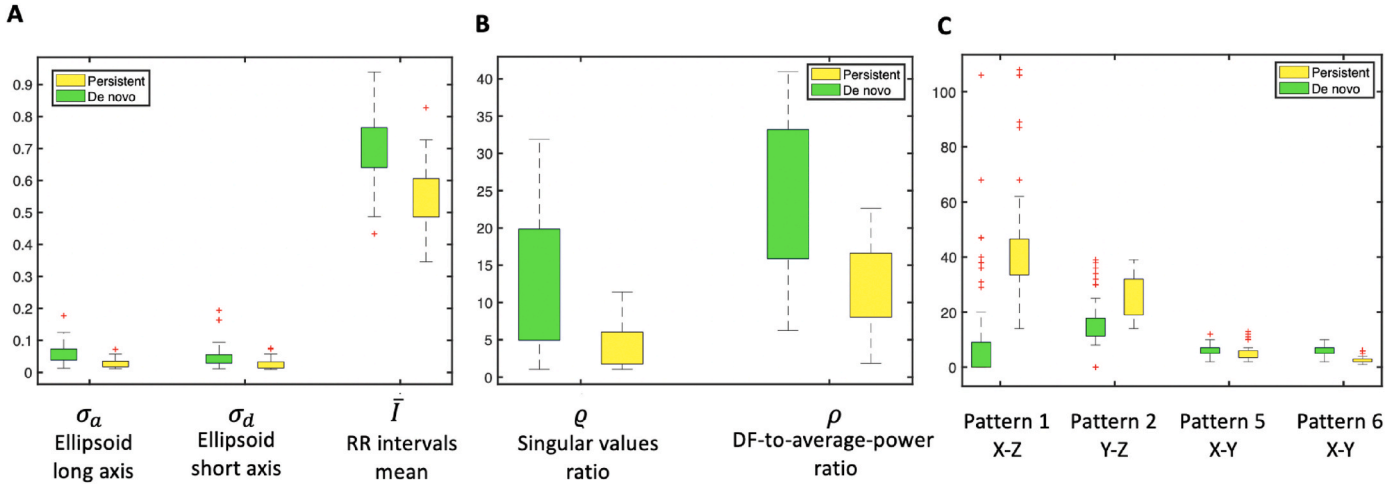

Fig. 12. Box plots of the selected features, A) Rhythm-based features (σ_a: ellipsoid long axis, σ_d: ellipsoid short axis and $\bar{I}$: RR intervals mean), B) Ratios (ϱ: singular values ratio Eq. (6), ρ: DF-to-average-power ratio Eq. (8)), C) VCG-based features.

(SEN), specificity (SPE), precision (PRE), and F_1 score. These metrics are defined as

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{10}$$

$$SEN = \frac{TP}{TP + FN} \times 100 \tag{11}$$

$$SPE = \frac{TN}{TN + FP} \times 100 \tag{12}$$

$$PRE = \frac{TP}{TP + FP} \times 100 \tag{13}$$

$$F_1 = 2\frac{PRE \times SEN}{PRE + SEN} \times 100 = \frac{2TP}{2TP + FP + FN} \times 100, \tag{14}$$

where TP and TN are the numbers of correctly detected and rejected cases in class one, respectively, and FP and FN are the numbers of wrongly detected and rejected cases in class one, respectively. We report the classification results for the random forest and SVM classifiers with kernel function RBF with scaling factor 1. Considering TP as the number of correctly detected cases of persistent AF, in Table 1 and Table 3, the performance of the SVM and RF classifier are shown, respectively. Considering TP as the number of correctly detected *de novo* POAF cases, in Table 2 and Table 4, the performance of the SVM and RF classifier are shown, respectively. To investigate the importance of each group of features, we trained the classifiers with each group of features separately (group 1: rhythm-based features, group 2: VCG-based features, group 3: frequency-based features). Then, we merged the groups with a combination of two out of three groups and finally, all three groups were used for training the classifiers. Looking at tbl1Tables 1 and 2tbl2, it is clear that all three groups and both classifiers are able to differentiate *de novo* POAF patients from persistent AF patients. The rhythm-based features and the combination of rhythm- and frequency-based features have the highest accuracy in the classification using one group and two groups, respectively. Moreover, the random forest has the highest accuracy and F_1 score. The classifier achieving the highest accuracy is the random forest classifier using all three groups of features, reaching an accuracy of over 89%.

Looking at Tables 1–4, Fig. 13, and Fig. 14, RF performs better than SVM. The performance of the SVMs depends on the kernel function. Although the RBF kernel is known to be a universal approximator [56], finding the optimal kernel parameters given our relatively limited dataset is a challenging problem. Note that in the current paper, our main goal was to demonstrate that classification is possible. Optimizing classification performance is out of the scope, however, we believe that SVM classification could be further improved by resolving the issues mentioned above.

In this work, we presented features which are able to differentiate between *de novo* POAF and persistent AF patients. Using these features might provide insights into the degree of the pathological damage in the tissue in AF patients. We studied explainable features from three different natures exploiting both morphological and rhythmic change information. The tracking of the progression of AF from the early stages (paroxysmal) to severe stages (persistent) is one of the most important actions in the treatment of AF. Since *de novo* POAF is the first stage

Table 1
The performance of the SVM classifier on the selected features, Group 1: Rhythm-based features, Group 2: VCG-based features, Group 3: Frequency-based features TP: Persistent AF.

Features	Metric	Acc	Sen	Spe	Pre	F_1
Group 1	**M**	71.84	75.89	67.48	71.77	73.33
	SD	3.28	8.13	10.01	6.55	3.73
Group 2	**M**	67.63	68.93	64.59	65.91	66.09
	SD	6.16	9.31	6.57	5.64	7.32
Group 3	**M**	68.91	71.25	66.12	67.82	68.89
	SD	4.38	4.12	9.32	8.01	7.56
Groups 1 & 2	**M**	72.91	74.77	71.40	70.50	72.14
	SD	4.61	8.14	7.08	6.17	3.22
Groups 1 & 3	**M**	74.14	64.04	82.90	76.33	66.28
	SD	5.45	6.23	6.92	5.46	5.62
Groups 2 & 3	**M**	72.99	73.98	71.51	70.76	71.94
	SD	5.68	7.20	7.26	7.16	5.79
All groups	**M**	79.00	81.38	75.96	78.30	79.69
	SD	5.65	4.82	8.95	7.44	7.12

Table 2
The performance of the SVM classifier on the selected features, Group 1: Rhythm-based features, Group 2: VCG-based features, Group 3: Frequency-based features TP: *De novo* POAF.

Features	Metric	Acc	Sen	Spe	Pre	F_1
Group 1	**M**	71.84	67.48	75.89	72.05	69.14
	SD	3.28	10.01	8.13	9.18	6.86
Group 2	**M**	67.63	64.59	68.93	69.77	65.95
	SD	6.16	6.57	9.31	6.17	7.81
Group 3	**M**	68.91	66.12	71.25	69.81	66.96
	SD	4.38	9.32	4.12	5.49	6.42
Groups 1 & 2	**M**	72.91	71.40	74.77	74.89	72.62
	SD	4.61	7.08	8.14	5.67	7.69
Groups 1 & 3	**M**	74.14	82.90	64.04	75.33	77.41
	SD	5.45	6.92	6.23	6.21	6.75
Groups 2 & 3	**M**	72.99	71.51	73.98	74.22	72.34
	SD	5.68	7.26	7.20	6.62	7.25
All groups	**M**	79.00	75.96	81.38	79.18	77.40
	SD	5.65	8.95	4.82	4.68	6.30

Table 3
The performance of the RF classifier on the selected features, Group 1: Rhythm-based features, Group 2: VCG-based features, Group 3: Frequency-based features TP: Persistent AF.

Features	Metric	Acc	Sen	Spe	Pre	F_1
Group 1	**M**	80.93	78.73	83.12	83.46	80.37
	SD	6.71	5.09	10.78	6.22	6.70
Group 2	**M**	77.01	70.21	82.58	80.95	72.66
	SD	5.05	6.18	4.98	5.22	5.79
Group 3	**M**	79.17	82.46	75.78	76.62	79.08
	SD	4.22	4.97	4.78	4.29	6.46
Groups 1 & 2	**M**	83.42	84.89	82.73	78.93	81.61
	SD	4.00	4.66	5.21	5.3	4.95
Groups 1 & 3	**M**	86.38	89.40	83.78	82.20	85.27
	SD	3.82	3.7	6.58	3.95	3.16
Groups 2 & 3	**M**	82.27	85.88	79.08	79.84	82.50
	SD	4.59	5.17	6.81	6.46	2.94
All groups	**M**	89.07	92.57	86.23	83.95	87.93
	SD	2.77	3.63	6.24	6.00	2.81

Table 4
The performance of the RF classifier on the selected features, Group 1: Rhythm-based features, Group 2: VCG-based features, Group 3: Frequency-based features TP: *De novo* POAF.

Features	Metric	Acc	Sen	Spe	Pre	F_1
Group 1	**M**	80.93	83.12	78.73	78.11	80.10
	SD	6.71	10.78	5.09	6.91	5.88
Group 2	**M**	77.01	82.58	70.21	75.91	77.55
	SD	5.05	4.98	6.18	6.12	5.91
Group 3	**M**	79.17	75.78	82.46	81.17	78.10
	SD	4.22	4.78	4.97	5.35	6.47
Groups 1 & 2	**M**	83.42	82.73	84.89	88.63	87.02
	SD	4.00	5.21	4.66	4.23	4.02
Groups 1 & 3	**M**	86.38	83.78	89.40	90.64	86.72
	SD	3.82	6.58	3.7	4.71	5.26
Groups 2 & 3	**M**	82.27	79.08	85.88	83.31	80.93
	SD	4.59	6.81	5.17	4.9	5
All groups	**M**	89.07	86.23	92.57	90.96	89.76
	SD	2.77	6.24	3.63	3.12	2.76

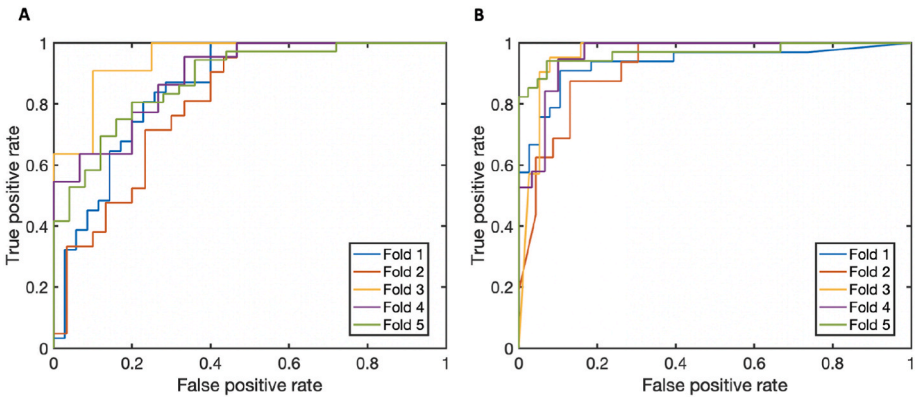

Fig. 13. Receiver operating characteristic (ROC) curve on the testing dataset, A) SVM, B) RF.

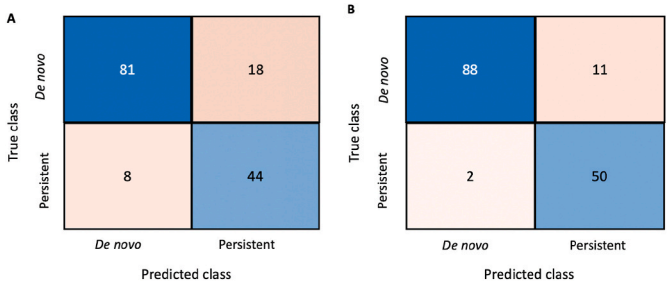

Fig. 14. Confusion matrix of the training dataset (the whole dataset), A) SVM, B) RF.

wherein a patient develops an AF episode, its early detection will help physicians to accurately monitor the progress of AF in the patient.

4. Discussion

4.1. Relation to literature

In general, detecting AF and doing the follow-up gives insights into the gradual progress of the pathological damage in the tissue. In the past, many studies have been done to find discriminators between AF and NSR. However, the stages of AF are often not included in these studies and the discrimination is based on the presence of an AF episode or an NSR episode. Self-terminating and sustained AF have been studied in recent works. Most of these studies focused on the dominant frequency of the AA, heart rate variability, sample entropy phase variations and fractal dimension [57–61]. However, they do not address the classification of *de novo* POAF and persistent AF. For instance, Refs. [59,60] compared self-terminating AF versus sustained AF, and e.g., Ref. [62] aimed to differentiate paroxysmal versus persistent AF. Notice that despite the fact that our problem overlaps with these classification tasks, it is not completely included in any of them. In addition, some of these works only compared AF episodes in short-length signals, e.g. Ref. [32], so the global features introduced in the current work cannot be directly compared with these.

In this work we considered both morphological (VCG-based and frequency-based features) and rhythmic (rhythm-based features) changes by extracting features from AF episodes. However, the performance of the rhythm-based and VCG-based features depends on the detection of the R peaks. We deal with this problem by extracting frequency-based features which are independent from the detection of R peaks. Therefore, it makes the algorithm more robust in case of missed-R peak detection. Related to the frequency-based features, Petrutiu et al. [57] calculated the dominant frequency of ECG signals by using template matching QRS-T cancellation and showed differences for paroxysmal and persistent AF. The dominant frequency in paroxysmal AF obtained was 5.2 ± 0.4 Hz while for persistent AF was 6.6 ± 0.6 Hz. Similarly, Chiarugi et al. [60] used the dominant frequency and average heart rate to discriminate between non-terminating and terminating AF episodes on 1-min ECG signals. These studies demonstrate the ability of the dominant frequency for differentiating paroxysmal/persistent AF or terminating/non-terminating AF. In the current study, we considered normalizing with the average power to take all involved frequency components into account.

4.2. Comparison with invasive wave mapping

Regardless of their discriminating capabilities, the dominant frequency and average power reveal pathological characteristics of both *de novo* POAF and persistent AF. That is, from a physiological point of view, the level of atrial fibrosis in paroxysmal AF patients is lower than that in persistent AF patients. So, paroxysmal AF patients tend to present a lower dominant frequency in comparison with persistent AF [32]. This observation is confirmed by our results. Using electrogram data, Allessie et al. [63] studied the wave maps and dissociation maps for the acute AF and longstanding AF obtained from the intra-operative mapping on the right atrium. It is noted that the number of waves entering the area of the mapping in acute AF is less than that in longstanding AF. The number of waves is also related to the boundaries of the conduction blocks and collisions. To compare our frequency-based features with the observation of wave mapping, we investigate the distribution of frequency-based features over the time. To do this, we divide the pre-processed signals into 60 s segments and calculate the index δ according to Eq. (5). Then, for AF episodes, we sort segments according to the δ from the highest to the lowest and select 20 segments with the highest δ. These are thus the most irregular segments. Then, we sub-divide each segment into 5 s non-overlapping frames and calculate the *DF* and ρ in each frame. Fig. 15 shows the histogram of the frequency-based features for 20 segments with the highest δ in a *de novo* POAF patient and a persistent AF patient. Comparing *DF* in a *de novo* POAF patient (Fig. 15A, red plot) and a persistent AF patient (Fig. 15A, purple plot), frame-to-frame variation of *DF* in *de novo* POAF is smaller than for persistent AF. It shows that by considering the whole spectrum, there are more frequency components in persistent AF that are activated. Looking at ρ in Fig. 15B, in the *de novo* POAF patient, the ρ is more concentrated than in persistent AF, which means that there is no direct relationship between *DF* and ρ in the persistent AF. Moreover, looking at ρ, there is less overlap between the *de novo* POAF and persistent AF compared to the *DF*, which shows that ρ could be a valuable discriminator for this classification. Therefore, combining these observations with wave mapping results, we hypothesize that there exists a relationship between the number of waves in the intra-operative mapping with the dominant frequency and power spectrum of the wavelet detail coefficients of the ECG signal. To test this hypothesis, we will extend this

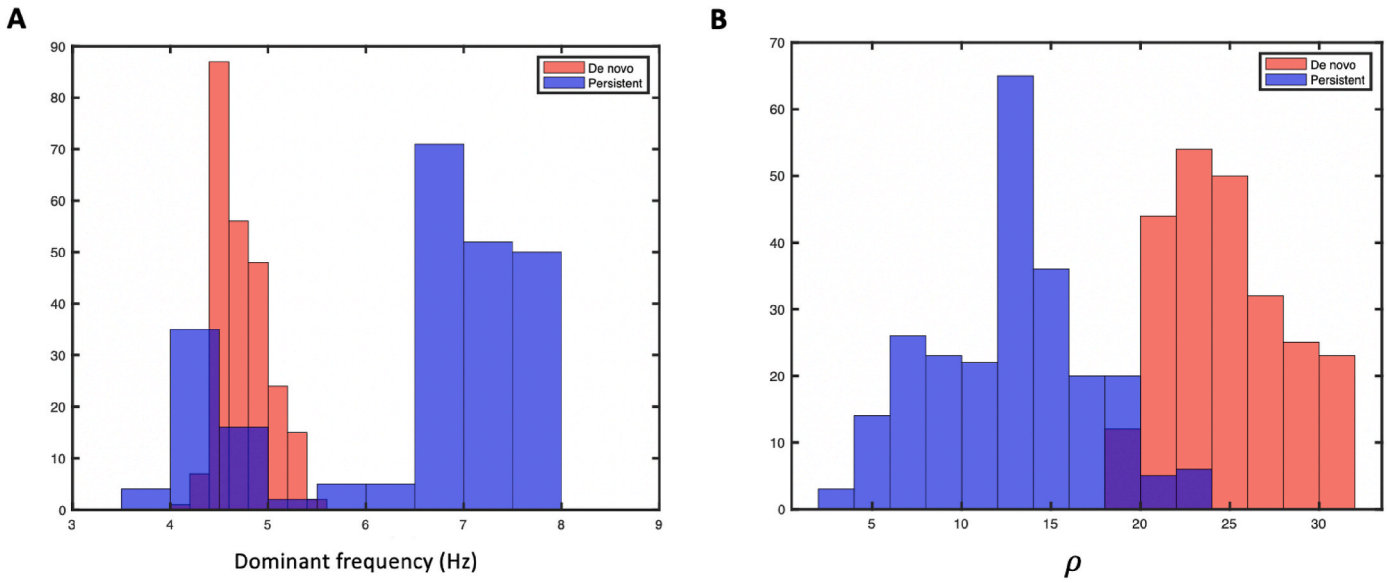

Fig. 15. Histogram of the frequency-based features for 20 segments with the highest δ, A) DF B) ρ.

in future work to instances where the electrogram and ECG signals are recorded simultaneously. The ρ will be calculated from leads *aVF* and III of the ECG (since V_1 is not available during an open-heart surgery) and then we will compare it with the results from the wave maps and dissociation maps obtained from the electrograms.

4.3. Limitation and future work

Although the performance of the proposed algorithm is good and it can assist clinicians in differentiating between two conditions, it has some limitations. First, we cannot really compare our work to other features/classification schemes as this is the first study on *de novo* POAF and persistent AF as far as we know. Moreover, the lack of telemetry data in the persistent AF cases limited our evaluation to a small group of patients. Results shown in the tables in Section 3 are based on a 5-fold cross-validation within our dataset of 151 patients. In order to investigate the robustness of our approach, future work will aim to validate the algorithm on an independent (preferably prospective and multi-center) dataset. Furthermore, body mass index (BMI) is an important factor in the amplitude of the recorded ECG signals, as well as for VCG signals. To be able to compare the surface of the QRS loops in *de novo* POAF and persistent AF effectively, BMI should also be considered. Finally, we focused on a specific frequency range in the frequency-based features to measure the dominant frequency of AA, however, in such a frequency range, the effects of VA are not completely removed. To improve the frequency-based features, QRS-T cancellation or AA extraction by source separation methods might lead to further improvement. Although our short-term goal is to improve the performance of the classification by finding the optimal classifier using a larger prospective dataset, there is a great interest in knowing if the features can unveil electro-pathological changes in AF patients. For this purpose, we will study the ECG signals and electrograms simultaneously to find the relationship between the wave maps and the frequency-based features. Finally, as predicting AF is necessary for increasing the success of treatments, we will concentrate on identifying features from the NSR episodes of the signal to predict AF. As *de novo* POAF is the most common complication after surgery, telemetry data from a large population of *de novo* POAF can be collected. Therefore, deep learning techniques could then be viable to identify features from the NSR episodes in a data-driven manner when combined with expert knowledge of the underlying physiological principles.

5. Conclusion

In this paper, we proposed a feature engineering machine learning-based classification approach for the detection of short- and long-term AF based on the extracted features from the multi-channel ECG signals. We introduced three groups of features (rhythm-, VCG- and frequency-based features) to cover both morphological and rhythmic changes from AF episodes. The performance of our method is evaluated by implementing RF and SVM classifiers in which RF could achieve 89.07% accuracy. Furthermore, the introduced features unveil the irregularity differences of the RR intervals and the morphological differences of the fibrillatory waves between *de novo* POAF and persistent AF patients. In other words, from the medical point of view, these features give important insight into the different nature of electropathology in *de novo* POAF vs persistent AF. Future work will focus on finding the relationship between the extracted features and the wave maps of intra-operative mapping.

Declaration of competing interest

None declared.

Acknowledgements

This research was funded in part by the Medical Delta Cardiac Arrhythmia Lab (CAL), the Netherlands.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2022.105270.

References

[1] Donald M. Lloyd-Jones, Thomas J. Wang, Eric P. Leip, Martin G. Larson, Daniel Levy, Ramachandran S. Vasan, Ralph B. D'Agostino, Joseph M. Massaro, Alexa Beiser, Philip A. Wolf, et al., Lifetime risk for development of atrial fibrillation: the Framingham heart study, Circulation 110 (9) (2004) 1042–1046.
[2] Leif Sörnmo, Pablo Laguna, Bioelectrical Signal Processing in Cardiac and Neurological Applications, ume 8, Academic Press, 2005.
[3] L. Samuel Wann, et al., 2011 ACCF/AHA/HRS focused update on the management of patients with atrial fibrillation, J. Am. Coll. Cardiol. 57 (2) (2011) 223–242.
[4] Valentin Fuster, et al., ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation, Eur. Heart J. 27 (16) (2006) 1979–2030.

[5] T January Craig, et al., 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation, J. Am. Coll. Cardiol. 74 (1) (2019) 104–132.
[6] Ebrahim Mansoor, De novo atrial fibrillation post cardiac surgery: the durban experience, Cardiovasc. J. Africa 25 (6) (2014) 282.
[7] Robert W. Rho, The management of atrial fibrillation after cardiac surgery, Heart 95 (5) (2009) 422–429.
[8] Malcolm Lewis, Dawood Parker, Clive Weston, Mark Bowes, Screening for atrial fibrillation: sensitivity and specificity of a new methodology, Br. J. Gen. Pract. 61 (582) (2011) 38–39.
[9] Allan J. Walkey, Syed K. Bashar, Md Billal Hossain, Eric Ding, Daniella Albuquerque, Michael Winter, Ki H. Chon, David D. McManus, Development and validation of an automated algorithm to detect atrial fibrillation within stored intensive care unit continuous electrocardiographic data: observational study, JMIR Cardio 5 (1) (2021), e18840.
[10] Daniele Marinucci, Agnese Sbrollini, Ilaria Marcantoni, Micaela Morettini, Cees A. Swenne, Laura Burattini, Artificial neural network for atrial fibrillation identification in portable devices, Sensors 20 (12) (2020) 3570.
[11] Jinho Park, Sangwook Lee, Moongu Jeon, Atrial fibrillation detection by heart rate variability in Poincare plot, Biomed. Eng. Online 8 (1) (2009) 38.
[12] J.P. Sepulveda-Suescun, J. Murillo-Escobar, R.D. Urda-Benitez, D.A. Orrego-Metaute, A. Orozco-Duque, Atrial fibrillation detection through heart rate variability using a machine learning approach and poincare plot features, in: VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th-28th, 2016, Springer, 2017, pp. 565–568.
[13] Chao Huang, Shuming Ye, Hang Chen, Dingli Li, Fangtian He, Yuewen Tu, A novel method for detection of the transition between atrial fibrillation and sinus rhythm, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 58 (4) (2010) 1113–1119.
[14] Douglas E. Lake, J Randall Moorman, Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices, Am. J. Physiol. Heart Circ. Physiol. 300 (1) (2011) H319–H325.
[15] Dhani Dharmaprani, Lukah Dykes, Andrew D. McGavigan, Pawel Kuklik, Kenneth Pope, Anand N. Ganesan, Information theory and atrial fibrillation (af): a review, Front. Physiol. 9 (2018) 957.
[16] Ayan Mukherjee, Anirban Dutta Choudhury, Shreyasi Datta, Chetanya Puri, Rohan Banerjee, Rituraj Singh, Arijit Ukil, Soma Bandyopadhyay, Arpan Pal, Sundeep Khandelwal, Detection of atrial fibrillation and other abnormal rhythms from ecg using a multi-layer classifier architecture, Physiol. Meas. 40 (5) (2019), 054006.
[17] Syed Khairul Bashar, Eric Ding, Daniella Albuquerque, Michael Winter, Sophia Binici, Allan J. Walkey, David D. McManus, Ki H. Chon, Atrial fibrillation detection in icu patients: a pilot study on mimic iii data, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 298–301.
[18] Jerzy Sacha, Interaction between heart rate and heart rate variability, Ann. Noninvasive Electrocardiol. 19 (3) (2014) 207–216.
[19] Alan Kennedy, Dewar D. Finlay, Daniel Guldenring, Raymond R. Bond, Kieran Moran, James McLaughlin, Automated detection of atrial fibrillation using rr intervals and multivariate-based classification, J. Electrocardiol. 49 (6) (2016) 871–876.
[20] Syed Khairul Bashar, Md Billal Hossain, Eric Ding, Allan J. Walkey, David D. McManus, Ki H. Chon, Atrial fibrillation detection during sepsis: study on mimic iii icu data, IEEE J. Biomed. Health Inform. 24 (11) (2020) 3124–3135.
[21] Saeed Babaeizadeh, Richard E. Gregg, Eric D. Helfenbein, James M. Lindauer, Sophia H. Zhou, Improvements in atrial fibrillation detection for real-time monitoring, J. Electrocardiol. 42 (6) (2009) 522–526.
[22] Yao Chen, Xiao Wang, Yonghan Jung, Vida Abedi, Ramin Zand, Marvi Bikak, Mohammad Adibuzzaman, Classification of short single-lead electrocardiograms (ecgs) for atrial fibrillation detection using piecewise linear spline and xgboost, Physiol. Meas. 39 (10) (2018) 104006.
[23] Yaru Yue, Chengdong Chen, Pengkun Liu, Ying Xing, Xiaoguang Zhou, Automatic detection of short-term atrial fibrillation segments based on frequency slice wavelet transform and machine learning techniques, Sensors 21 (16) (2021) 5302.
[24] Dakun Lai, Yuxiang Bu, Ye Su, Xinshu Zhang, Chang-Sheng Ma, Non-standardized patch-based ecg lead together with deep learning based algorithm for automatic screening of atrial fibrillation, IEEE J. Biomed. Health Inform. 24 (6) (2020) 1569–1578.
[25] Jibin Wang, Ping Wang, Suping Wang, Automated detection of atrial fibrillation in ecg signals based on wavelet packet transform and correlation function of random process, Biomed. Signal Process Control 55 (2020) 101662.
[26] Iben H. Bruun, Semira M.S. Hissabu, Erik S. Poulsen, Sadasivan Puthusserypady, Automatic atrial fibrillation detection: a novel approach using discrete wavelet transform and heart rate variability, in: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC), 2017, pp. 3981–3984.
[27] Ehab Essa, Xianghua Xie, An ensemble of deep learning-based multi-model for ecg heartbeats arrhythmia classification, IEEE Access 9 (2021) 103452–103464.
[28] Saurav Mandal, Nabanita Sinha, Prediction of atrial fibrillation based on nonlinear modeling of heart rate variability signal and svm classifier, Res. Biomed. Eng. (2021) 1–12.
[29] Gerald Hirsch, Søren H. Jensen, Erik S. Poulsen, Sadasivan Puthusserypady, Atrial fibrillation detection using heart rate variability and atrial activity: a hybrid approach, Expert Syst. Appl. 169 (2021) 114452.
[30] Y-Y Jo, Younghoon Cho, Soo Youn Lee, Joon-myoung Kwon, Kyung-Hee Kim, Ki-Hyun Jeon, Soohyun Cho, Jinsik Park, Byung-Hee Oh, et al., Explainable artificial intelligence to detect atrial fibrillation using electrocardiogram, Int. J. Cardiol. 328 (2021) 104–110.
[31] Xianjie Chen, Zhaoyun Cheng, Sheng Wang, Guoqing Lu, Gaojun Xv, Qianjin Liu, Xiliang Zhu, Atrial fibrillation detection based on multi-feature extraction and convolutional neural network for processing ecg signals, Comput. Methods Progr. Biomed. 202 (2021) 106009.
[32] Nuria Ortigosa, Antonio Galbis, Carmen Fernández, Óscar Cano, Gabor frames for classification of paroxysmal and persistent atrial fibrillation episodes, Med. Eng. Phys. 39 (2017) 31–37.
[33] Kligfield Paul, et al., Recommendations for the standardization and interpretation of the electrocardiogram: part i: the electrocardiogram and its technology, J. Am. Coll. Cardiol. 49 (10) (2007) 1109–1127.
[34] Jiří Kozumplík, Ivo Provazník, Fast time-varying linear filters for suppression of baseline drift in electrocardiographic signals, Biomed. Eng. Online 16 (1) (2017) 1–16.
[35] Kang-Ming Chang, Arrhythmia ECG noise reduction by ensemble empirical mode decomposition, Sensors 10 (6) (2010) 6063–6080.
[36] Juan Pablo Martínez, Rute Almeida, Salvador Olmos, Ana Paula Rocha, Pablo Laguna, A wavelet-based ECG delineator: evaluation on standard databases, IEEE Trans. Biomed. Eng. 51 (4) (2004) 570–581.
[37] M. Malik, Heart rate variability: standards of measurement, physiological interpretation, and clinical use, Ann. Noninvasive Electrocardiol. 1 (2) (1996) 151–181.
[38] M. Brennan, M. Palaniswami, P. Kamen, Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability? IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 48 (11) (2001) 1342–1347.
[39] M.H. Hayes, Statistical Digital Signal Processing and Modeling, John Wiley and Sons, New York, 1996, ISBN 0-471 59431-8.
[40] H. Akaike, Fitting autoregressive models for prediction, Ann. Inst. Stat. Math. 21 (1) (1969) 243–247.
[41] Ernest Frank, An accurate, clinically practical system for spatial vectorcardiography, Circulation 13 (5) (1956) 737–749.
[42] Eedara Prabhakararao, Samarendra Dandapat, Automated detection of posterior myocardial infarction from VCG signals using stationary wavelet transform based features, IEEE Sens. Lett. 4 (6) (2020) 1–4.
[43] Timo Ojala, Matti Pietikäinen, David Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recogn. 29 (1) (1996) 51–59.
[44] Timo Ahonen, Abdenour Hadid, Matti Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.
[45] Caifeng Shan, Shaogang Gong, Peter W. McOwan, Facial expression recognition based on local binary patterns: a comprehensive study, Image Vis Comput. 27 (6) (2009) 803–816.
[46] Hanie Moghaddasi, Saeed Nourian, Automatic assessment of mitral regurgitation severity based on extensive textural features on 2D echocardiography videos, Comput. Biol. Med. 73 (2016) 47–55.
[47] Arun Balodi, R.S. Anand, M.L. Dewal, Anurag Rawat, Severity analysis of mitral regurgitation using discrete wavelet transform, IETE J. Res. (2020) 1–11.
[48] Timo Ojala, Matti Pietikainen, Topi Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.
[49] Shadnaz Asgari, Alireza Mehrnia, Maryam Moussavi, Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine, Comput. Biol. Med. 60 (2015) 132–142.
[50] Nekane Larburu, T. Lopetegi, I. Romero, Comparative study of algorithms for atrial fibrillation detection, in: 2011 Computing in Cardiology, IEEE, 2011, pp. 265–268.
[51] Binwei Weng, John J. Wang, Francis Michaud, Manuel Blanco-Velasco, Atrial fibrillation detection using stationary wavelet transform analysis, in: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2008, pp. 1128–1131.
[52] J.J. Rieta, J. Millet-Roig, V. Zarzoso, F. Castells, C. Sanchez, R. Garcia-Civera, S. Morell, Atrial fibrillation, atrial flutter and normal sinus rhythm discrimination by means of blind source separation and spectral parameters extraction, in: Computers in Cardiology, IEEE, 2002, pp. 25–28.
[53] Marko Robnik-Šikonja, Igor Kononenko, Theoretical and empirical analysis of ReliefF and RReliefF, Mach. Learn. 53 (1–2) (2003) 23–69.
[54] Tin Kam Ho, A data complexity analysis of comparative advantages of decision forest constructors, Pattern Anal. Appl. 5 (2) (2002) 102–112.
[55] Alberto Fernández, Salvador García, Mikel Galar, Ronaldo C. Prati, Bartosz Krawczyk, Francisco Herrera, Learning from Imbalanced Data Sets, ume 10, Springer, 2018.
[56] Junping Wang, Quanshi Chen, Yong Chen, Rbf kernel based support vector machine with universal approximation and its application, in: International Symposium on Neural Networks, Springer, 2004, pp. 512–517.
[57] Simona Petrutiu, Alan V. Sahakian, Steven Swiryn, Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans, Europace 9 (7) (2007) 466–470.
[58] Nuria Ortigosa, Carmen Fernández, Antonio Galbis, Óscar Cano, Phase information of time-frequency transforms as a key feature for classification of atrial fibrillation episodes, Physiol. Meas. 36 (3) (2015) 409.
[59] Raúl Alcaraz, José J. Rieta, Sample entropy of the main atrial wave predicts spontaneous termination of paroxysmal atrial fibrillation, Med. Eng. Phys. 31 (8) (2009) 917–922.
[60] Franco Chiarugi, Maurizio Varanini, Federico Cantini, Fabrizio Conforti, Giorgos Vrouchos, Noninvasive ECG as a tool for predicting termination of

paroxysmal atrial fibrillation, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 54 (8) (2007) 1399–1406.
[61] Bahareh Safarbali, Seyed Mohammad Reza Hashemi Golpayegani, Nonlinear dynamic approaches to identify atrial fibrillation progression based on topological methods, Biomed. Signal Process Control 53 (2019) 101563.
[62] Raúl Alcaraz, José Joaquín Rieta, The application of nonlinear metrics to assess organization differences in short recordings of paroxysmal and persistent atrial fibrillation, Physiol. Meas. 31 (1) (2009) 115.
[63] Maurits A. Allessie, Natasja MS. de Groot, Richard PM. Houben, Schotten Ulrich, Eric Boersma, Joep L. Smeets, Harry J. Crijns, Electropathological substrate of long-standing persistent atrial fibrillation in patients with structural heart disease: longitudinal dissociation, Circul.: Arrhythmia Electrophysiol. 3 (6) (2010) 606–615.