Atrial Fibrillation Stratification Via Super-Resolution Analysis of Fibrillatory Waves

Saumitra Mishra, Sreehari Rammohan, Khalid Z Rajab, Gurpreet Dhillon, Pier Lambiase, Ross Hunter, Elaine Chew

Atrial Fibrillation Stratification Via Super-Resolution Analysis of Fibrillatory Waves

Saumitra Mishra[1], Sreehari Rammohan[2], Khalid Z. Rajab[1], Gurpreet Dhillon[1,3], Pier Lambiase[3,4], Ross Hunter[1,3], Elaine Chew[5]

[1]Queen Mary University of London, London, UK [2]Cupertino High School, California, USA [3]Barts Heart Centre, St. Bartholomew's Hospital, London, UK [4]University College London Institute of Cardiovascular Science, London, UK [5]CNRS−UMR9912 / STMS Laboratory, Paris, France

Abstract

We use the Filter Diagonalization Method (FDM), a harmonic inversion technique, to extract f-wave features in electrocardiographic (ECG) traces for atrial fibrillation (AF) stratification. The FDM detects f-wave frequencies and amplitudes at frame sizes of 0.15 seconds. We demonstrate our method on a dataset comprising of ECG recordings from 23 patients (61.65 ± 11.63 *years,* 78.26% *male) before cryoablation; 2 paroxysmal AF, 16 early persistent AF (<12 months duration), and 4 longstanding persistent AF (>12 months duration). Moreover, some of these patients received adenosine to enhance their RR intervals before ablation. Our method extracts features from FDM outputs to train statistical machine learning classifiers. Ten-fold cross-validation demonstrates that the Random Forest and Decision Tree models performed best for the pre-ablation without and with adenosine datasets, with accuracy* $60.89 \pm 0.31\%$ *and* $59.58\% \pm 0.04\%$*, respectively. While the results are modest, they demonstrate that f-wave features can be used for AF stratification. The accuracies are similar for the two tests, slightly better for the case without adenosine, showing that the FDM can successfully model short f-waves without the need to concatenate f-wave sequences or adenosine to elongate RR intervals.*

1. Introduction

Atrial fibrillation (AF) is a fast growing global epidemic [1]. Treatment decisions regarding the feasibility of restoring sinus rhythm are based on assessments of left atrial size and the chronicity of AF, but these factors give only an approximate assessment of disease progression and are poor discriminators of who will respond to a rhythm control strategy. This clinical classification of AF is crude, with little bearing on symptom severity, progression risk, and treatment success.

AF is characterised by fibrillatory waves (f-waves) in the electrocardiographic (ECG) signal that alter in shape, size, and organization as the disease progresses and the atria undergo structural remodelling. Properties of f-waves have been used for risk stratification. Bollman et al. [2] showed f-wave frequency to be strongly correlated to the atrial defibrillation threshold for persistent AF (PeAF) patients undergoing cardioversion. In Cheng et al. [3], f-wave amplitude was shown to predict persistent AF recurrence in patients who underwent catheter ablation. Zeemering et al. [4] showed higher dominant frequency and higher f-wave amplitude to be associated with increased risk of progression to PeAF. Alcaraz and Rieta [5] used Wavelet Entropy to characterise the f-wave disorder degree to predict successfully spontaneous termination of paroxysmal AF (PAF) and cardioversion outcome in PeAF patients.

Here, we propose and test the use of the Filter Diagonalization Method (FDM) to extract f-wave features in ECG traces for AF stratification. The FDM is an effective and highly efficient harmonic inversion technique that allows for super-resolution extraction of harmonic frequencies and characteristics over a very short window, in comparison to the period of the signal. In [6], the FDM was successfully used to detect musical vibratos, which typically range between 4–8 Hz, and characterize their features. F-waves exhibit similar undulatory behaviors and range between 3–12 Hz. An FFT-based method would require multiple time periods in order to extract the harmonic features with sufficient resolution, and f-wave features could risk being averaged out and lost in the noise floor. Features extracted using the FDM are used to train statistical machine learning classifiers to stratify patients into the labeled subgroups. In the present experiments, ECG data is drawn from PeAF (early and longstanding) and PAF patients prior to ablation, with some patients being administered adenosine to elongate their RR intervals. Since there are presently no established categories of AF which are particularly useful or reflect disease progression, to demonstrate biological plausibility of categorising

patients using this technique, we compare the results to the conventional categorizations of PAF, Early PeAF (E-PeAF), and longstanding PeAF (LS-PeAF).

2. The Filter Diagonalization Method

In the FDM, the idea is to decompose a signal into a Fourier basis. This is done by matching it to a filter composed of a grid of frequencies points (which serve as initial estimates of the harmonics), diagonalizing to place it into a solvable (and computational efficient) form, and then obtaining the relevant harmonic characteristics through solution of an eigenequation.

A signal, such as an ECG waveform, can be expressed as a sum of complex exponentials,

$$g_n = g(n\tau) = \sum_{k=0}^{K} d_k e^{-in\tau\omega_k} \tag{1}$$

where $n = 0, 1, 2, ..., N$ is an integer, and τ is the measurement sampling time and ω_k is the frequency. A key novelty is the recognition that this form of a time-series waveform can be associated with an autocorrelation function , and solving for the spectral features of g is equivalent to diagonalizing the evolution operator, $\hat{U} = e^{-i\tau\hat{\Omega}}$ [7].

A matrix operator $\mathbf{U}$ is defined, in a similar manner, and is equivalent to a two-dimensional Fourier transform,

$$\mathbf{U}^{(p)} = \sum_{n=0}^{N}\sum_{n'=0}^{N} e^{in\phi}e^{in'\phi'} g_0\left(n + n' + p\right). \tag{2}$$

To evaluate this operator, a uniform grid of frequency search components is defined in the vicinity of the spectral harmonic components, $\phi_k = -2\pi\left(k\Delta h + h_{min}\right)\tau$, where (h_{min}, h_{max}) defines the search range for the harmonic frequencies, and $\Delta h = \frac{1}{K}(h_{max} - h_{min})$. The Nyquist criterion sets the search resolution of the grid, $K = \left(h_{max} - h_{min}\right) N\tau/2 - 1$.

Using the definition of the evolutionary operator, solve for the eigenvalue equation:

$$\mathbf{U^{(1)}B_k} = \mathbf{u_k U^{(1)} B_k}. \tag{3}$$

Each of the k solutions corresponds to a spectral harmonic, such that $h_k = \frac{i}{2\pi\tau}\ln(u_k)$. Note that the real part of h_k corresponds to the frequency component, and the imaginary part represents the decay of the wave. The amplitude of this component of the signal is found using the eigenvectors $\mathbf{B}$, such that $d_k = \left(\sum_{n=0}^{N} g(n) B_{nk}\right)^2$, where $|d_k|$ is the amplitude of the signal, and the phase is the time delay. Note that, because the size of the problem is proportional to the number of harmonics, K, rather than the number of time samples, $N >> K$, the computational burden is low.

Table 1. Dataset summary. Pre and PreAdenosine are two types of ECG excerpts used for the experiments. For each AF category C, N_p^C, N_r^C and N_{TQ}^C refer to the number of patients, recordings and TQ segments, respectively.

AF	Pre			PreAdenosine		
	N_p^{Pre}	N_r^{Pre}	N_{TQ}^{Pre}	N_p^{PreA}	N_r^{PreA}	N_{TQ}^{PreA}
PAF	2	9	4461	2	9	4461
E-PeAF	16	16	5108	7	7	2225
LS-PeAF	4	4	1350	1	1	474

In summary, the FDM is a highly efficient technique—computational complexity is a function of number of harmonics—that can accurately resolve over a small number of waveform measurements. This is well-suited to extracting fine features exhibited over a short period of time, over which FFT-based methods would not be suitable.

3. Experiments

This section describes the experiments that demonstrate the effectiveness of the FDM for AF stratification.

3.1. Dataset

The experiment draws ECG data from 23 patients, 18 men and 5 women aged 61.65 ± 11.63 years, consented for the Barts BioResource database who are participating in a larger CardioInsight (ECGI) study (NCT03394404). 17 of these are E-PeAF patients (AF duration < 12 months), 4 are LS-PeAF patients (AF duration > 12 months), and two are PAF patients. Of these, 7 E-PeAF, and one LS-PeAF patient were administered adenosine to enhance their RR intervals for analysis before ablation.

All patients underwent cryo-ablation. However, the current experiment only uses data preceding the ablation. We extract 10 minute excerpts from the ECG (sampling rate 2 kHz) recorded by the digital recording system (Bard LabSystem Pro, Boston Scientific), and label them 'Pre'. For patients who received adenosine, we also excerpt ECG data after adenosine and label that 'PreAdenosine'(PreA). All except one patient with PreA segments also had Pre segments. Table 1 shows the dataset distribution for the experiments.

3.2. Feature Extraction

Our method extracts features from ECG excerpts in two steps. We first extract TQ segments—the segments from the end of the T wave to the start of the next QRS complex. Then, we apply the FDM to each frame in a TQ segment to generate one feature vector per TQ segment. Figure 1 depicts the feature extraction pipeline.

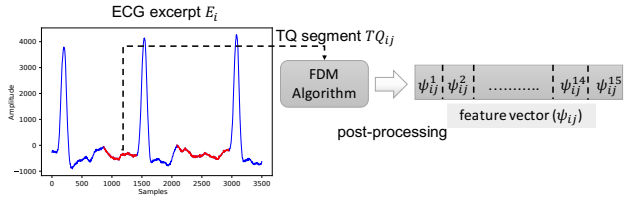

Figure 1. Functional block diagram of the feature extraction pipeline. Red segments depict TQ segments. TQ_{ij} is the j_{th} TQ segment of the ith ECG excerpt. The FDM returns per frame frequency and amplitude information, post-processing which generates the feature vector $[\psi_{ij}]$.

Our method extracts TQ segments by first finding R peaks using a computed minimum amplitude threshold and a minimum distance between R peaks, and then splicing the signal using known average times for QRS and TQ complexes[1]. Table 1 shows the number of TQ segments per AF category for Pre and PreA excerpt types.

We apply the FDM to overlapping frames of each TQ segment by using a 150 ms window with 10 ms hop size. The FDM returns a set of frequencies and corresponding magnitudes for each frame, t, that we use to locate the maximum magnitude, a_t^{max}, and its corresponding frequency, f_t^{max}, per frame. Over all frames, we select the set of valid frequencies, $F_{val} = \{f_t | f_t \in [3, 12]\}$, those in the f-wave range. We then construct the 15-dimensional feature vector, $[\psi_{ij}]$, for the j-th TQ segment in the ith-ECG excerpt:

1. Features 1-3: f_t^{max} mean, median, standard deviation
2. Features 4-6: $2a_t^{max}$ mean, median, standard deviation
3. Features 7-15: Normalised F_{val} distribution

Thus, we create a set of feature vectors, $\Psi = [\psi_i]$, and matching labels, $\Gamma = [\gamma_i]$, for $i = 1 \ldots N_r^C$. Each ψ_i represents a 15-dimensional feature vector, $[\psi_{ij}]$, and each γ_i points to an AF category, {PAF, E-PeAF, LS-PeAF}.

3.3. Model Training and Evaluation

We train Decision Tree (DT) [8] and Random Forest (RF) [9] classifiers for Pre and for PreA excerpts. Table 1 shows that both the Pre and PreA subsets are highly imbalanced. The number of TQ segments in LS-PeAF is much smaller than that in the other two AF categories. Training a model on such a dataset may bias the predictions to the dominant AF categories. To avoid this, from the Pre and PreA datasets, we create a data subset by randomly selecting about 500 TQ segments per AF category to form the $\Psi_{sampled}^{Pre}$ and $\Psi_{sampled}^{PreA}$ datasets with 1500 and 1474 features, respectively.

[1] *https://www.nottingham.ac.uk/nursingpractice/resources/cardiology/function/normal_duration.php* (accessed 15 July 2019)

Table 2. Average classification accuracy (in %) and standard deviation for the best decision tree and random forest models for each ECG excerpt type.

Classifier	ECG excerpt type	
	Pre	PreAdenosine
Random baseline	33.33	33.33
Decision tree	59.33($\pm$0.038)	**59.58($\pm$0.040)**
Random forest	**60.89($\pm$0.311)**	58.22($\pm$0.035)

We perform two types of evaluations per classifier—cross-validation and holdout dataset-based evaluation. To do this, we split each of the two sampled datasets into two subsets of unequal sizes. For example, for the Pre subset, we generate the cross-validation Ψ_{CV}^{Pre} and holdout evaluation Ψ_{Eval}^{Pre} subsets by randomly splitting $\Psi_{sampled}^{Pre}$ into 90% and 10% subsets, respectively. Similarly, we create Ψ_{CV}^{PreA} and Ψ_{Eval}^{PreA} datasets from $\Psi_{sampled}^{PreA}$.

In the first evaluation, we perform 10-fold cross-validation over the cross-validation subsets and report the average classification accuracy for each trained model for the Pre and PreA subsets. We select the best model by performing a random search over the hyperparameter space and choosing the model with the best average classification accuracy over 10 folds. Table 2 reports the results of 10-fold cross-validation for the best DT and RF models for the Pre and PreA subsets. The evaluation results demonstrate that all the classification models perform much better than a random baseline (33%). Moreover, RF and DT models perform best for the Pre and PreA subsets, respectively. The results also demonstrate that the average classification accuracy of the best models for both the recording categories is nearly the same—the difference is only around 1%. This demonstrates that the FDM works equally well for short TQ segments (in Pre) as for longer ones (in PreAdenosine).

In the second evaluation, we select the best model, one with the highest average classification category over the cross-validation subsets, for the Pre and PreA subsets, and evaluate its performance on the holdout evaluation subsets. The holdout subsets are withheld during training, hence evaluation on them provides an unbiased estimate of the models' generalizability. Table 3 presents the evaluation results for the holdout subsets. Figure 2 plots the confusion matrix for each model from Table 3.

4. Conclusions and Future Work

We proposed and demonstrated the effectiveness of FDM-based features for AF stratification. The FDM effectively extracted key spectral and amplitude information from short-duration f-waves which, when modelled with machine learning classifiers, resulted in classification ac-

Table 3. Best model performance statistics for holdout datasets.'Overall' reports metrics for all AF categories; the metrics per AF category reports how well the model predicts presence/absence of the AF category, giving the accuracy (Acc), precision (P), recall (R), and F1-score (F1).

Category	Best Model	AF category	Acc	P	R	F1
Pre	Random Forest	Overall	57%	0.59	0.57	0.57
		PAF		0.48	0.63	0.55
		E-PeAF		0.60	0.48	0.54
		LS-PeAF		0.66	0.62	0.64
PreA	Decision Tree	Overall	56%	0.60	0.56	0.56
		PAF		0.67	0.71	0.69
		E-PeAF		0.41	0.64	0.50
		LS-PeAF		0.67	0.39	0.49

curacies around 60%, which is well above chance.

Although we have compared the output categories to conventional classifications (for lack of an established alternative) it is worth noting that this serves as a poor ground truth as it is a poor reflector of disease progression. Furthermore, some E-PeAF patients were previously cardioverted, and might be better categorized between PAF and E-PeAF. This presentation is intended only to demonstrate biological plausibility of the results.

We demonstrated that the best models for both Pre and PreAdenosine excerpts have nearly the same performance. This highlights the fact that the FDM can effectively extract discriminative features from short-duration f-waves, avoiding the need for adenosine, which can alter f-wave behaviour. This makes the system more straightforward to apply, potentially through a standard ECG.

Future work includes extending the experiments to a larger, more balanced dataset and considering other FDM generated features such as time delay. We will seek to improve the classification accuracy by using other classifiers and to analyse the behaviour of the best performing classification models. It is hoped that the system will ultimately stratify patients in a more meaningful way than conventional clinical classifications. Once finished, we plan to test the predictive power of the system to determine the likelihood of success of a rhythm control strategy, which could then be readily deployed through an ECG.

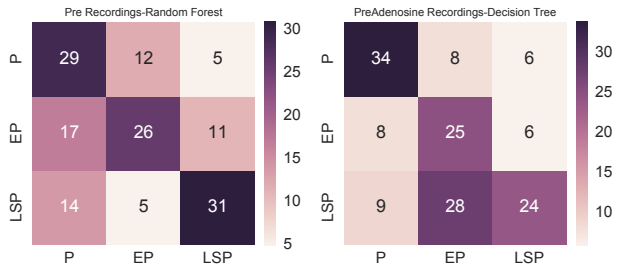

Figure 2. Confusion matrices for the best model for Pre and PreAdenosine (P=PAF, EP=E-PeAF, LSP=LS-PeAF). The number in each box gives the number of TQ segments.

Acknowledgements

Clinical ECG data were taken from a trial (NCT03394404) sponsored by Medtronic. This work was supported in part by a QMI PoC small grant and by the ERC under the European Union's Horizon 2020 research and innovation programme (Grant agreement No.788960).

References

[1] Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, Gillum RF, Kim YH, McAnulty JHJ, Zheng ZJ, Forouzanfar MH, Naghavi M, Mensah GA, Ezzati M, Murray CJL. Worldwide epidemiology of AF: A global burden of dis. 2010 study. Circulation 2013;129(8):837–847.

[2] Bollmann A, Mende M, Neugebauer A, Pfeiffer D. Atrial fibrillatory frequency predicts atrial defibrillation threshold and early arrhythmia recurrence in patients undergoing internal cardioversion of persistent atrial fibrillation. Pacing Clin Electrophysiol 2002;25(8):1179–1184.

[3] Cheng Z, Deng H, Cheng K, Chen T, Gao P, Yu M, Fang Q. The amplitude of fibrillatory waves on leads avf and v1 predicting the recurrence of persistent atrial fibrillation patients who underwent catheter ablation. Ann Noninv Electrocard 2013;18(4):352–358.

[4] Zeemering S, Lankveld T, Bonizzi P, Limantoro I, Bekkers S, Crijns H, Schotten U. The electrocardiogram as a predictor of successful pharmacological cardioversion and progression of AF. EP Europace 2018;20(7):e96–e104.

[5] Alcaraz R, Rieta JJ. Application of wavelet entropy to predict atrial fibrillation progression from the surface ecg. Comp Math Methods in Med 2012;245213.

[6] Yang L, Rajab KZ, Chew E. Filter diagonalisation method for music signal analysis: Frame-wise vibrato detection and estimation. J Maths Music Mar. 2017;11(1):42–60.

[7] Wall M, Neuhauser D. Extraction, through filter-diagonalization, of general quantum eigenvalues or classical normal mode freqs from a small no. of residues or a short-time segment of a signal. i. theory and applic to a quantum-dynamics model. J Chem Phy 1995;102(20):8011–8022.

[8] Breiman L, J F, Stone C, Olshen R. Classification and Regression Trees. CRC Press, 1984.

[9] Breiman L. Random forests. Machine Learning 2001; 45(1):5–32.

Address for correspondence:

Elaine Chew
IRCAM, 1 place Igor-Stravinsky, 75004 Paris, France.
elaine.chew@ircam.fr