Biomedical Signal Processing and Control 38 (2017) 182–190

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Feature subset selection and classification of intracardiac electrograms during atrial fibrillation

S.I. Duque [a,*], A. Orozco-Duque [a,b], V. Kremen [c], D. Novak [d], C. Tobón [e], J. Bustamante [a]

[a] Bioengineering Center, Universidad Pontificia Bolivariana, Medellín, Colombia
[b] GI²B, Instituto Tecnológico Metropolitano, Medellín, Colombia
[c] Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague, Prague, Czech Republic
[d] Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic
[e] MATBIOM, Universidad de Medellín, Medellín, Colombia

ARTICLE INFO

Article history:
Received 17 May 2016
Received in revised form 22 May 2017
Accepted 4 June 2017
Available online 23 June 2017

Keywords:
Atrial fibrillation
Electroanatomical mapping
Fractionated electrograms
K-NN classifier
Rotor

ABSTRACT

Several approaches have been adopted for the identification of arrhythmogenic sources maintaining atrial fibrillation (AF). In this paper, we propose a classifier that discriminates between four classes of atrial electrogram (EGM). We delved into the relation between levels of fractionation in EGM signals and the fibrillation substrates in simulated episodes of chronic AF. Several feature extraction methods were used to calculate 92 features from 429 real EGM records acquired during radiofrequency ablation of chronic AF. We selected the optimal subset of features by using a genetic algorithm, followed by *K*-nearest neighbors (K-NN) classification into four levels of fractionation. Sensitivity of 0.90 and specificity of 0.97 were achieved. Subsequently, the results of the classification were extrapolated to signals of a 3D human atria model and a 2D model of atrial tissue. The 3D model simulated an episode of AF maintained by a rotor in the posterior wall of the left atrium and the 2D model simulated an AF episode with one stable rotor. We used the K-NN classifier trained on a given set of real EGM signals to detect a specific class of signals presenting the highest level of fractionation located near the rotor's vortex. This method needs to be tested on real clinical data to provide evidence that it can support ablation therapy procedures.

1. Introduction

Atrial fibrillation (AF) is the most common tachyarrhythmia encountered in clinical practice. AF Patients have twice the mortality rate of the healthy population their incidence of stroke is six times higher [1].

AF diagnosis is performed by analyzing electrocardiographic (ECG) signals and intracardiac electrograms (EGM). They are recordings of cardiac potentials detected by electrodes placed in direc contact with the endocardium [2]. AF treatment aims to restore the sinus rhythm and atrial function. Although medication is used to control the heart rate and to prevent coagulation, some patients continue to experience symptoms despite medical therapy. Surgical radiofrequency ablation techniques, such as electrical pulmonary veins isolation (PVI), have therefore been used as treatment for symptomatic chronic AF. However, these techniques need to be improved [3].

Some new ablation techniques propose strategies for identifying arrhythmogenic sources by mapping EGM features. Characteristics such as areas of activation [4], dominant frequency [5] and areas presenting complex fractionated atrial electrograms (CFAE) are considered potential ablation sites [6]. The approach mentioned in [6] has been implemented in commercial equipment [7]. According to Nademanee, CFAE could be used as target sites for ablation and are formally defined as follows: (1) atrial electrograms that have fractionated activations composed of two or more deflections and/or perturbation of the baseline with continuous deflection of a prolonged activation complex over a 10 s recording period; (2) atrial electrograms with a very short cycle length ($\leq$120 ms) over a 10 s recording period. However, several studies have shown results conflicting with those reported by Nademanee et al. [8,9]. The inexact and vague definition of CFAE by Nademanee and colleagues might not be enough to differentiate the levels of fractionation and the morphologies in the EGMs. Therefore, according to their definition of CFAE, fractionated EGMs may not be related to arrhythmogenic substrates [10]. Moreover, semi-automated bipolar CFAE algorithms implemented in commercial equipment showed poor correlation with each other [11]. These

* Corresponding author.
E-mail address: sara.duquevallejo@gmail.com (S.I. Duque).

http://dx.doi.org/10.1016/j.bspc.2017.06.005
1746-8094/

considerations leads to subjective interpretations by the specialist when deciding which fragmented EGMs should be targeted for successful ablation. In consequence, AF ablation remains limited to the isolation of the pulmonary veins. When ablation is guided by electrogram, success greatly depends on the experience of the electrophysiologist [12,13].

As a consequence, several approaches have been proposed to distinguish levels of fractionation in the EGM. One of the first classification systems of EGM signals was proposed by Wells et al. [14]. They classified four types of atrial EGMs based on the presence of activations in the signal and the nature of their baseline. Schilling et al. [15] created a classification system which can be considered as a modification of the Wells' criteria. It is composed of 4 groups, it uses a new criterion, and it applies a Fuzzy Decision Tree. The four groups are: *non-fractionated with high frequency; fractionated with periodic activity; unstable electrograms with a mixture of periodic fractionated and periodic non-fractionated atrial electrograms*; and *continuous activity*. Studies of different levels of fractionation identify morphologies in the signal, thus assisting the localization of critical sites in AF.

Experimental evidence suggests that certain types of AF are the result of a phenomenon known as a rotor: the continuous circulation of an electrical impulse around a singular point [16]. Some studies have presented evidence linking the vortex of the rotor with a high level of fractionation in the EGM [17–19]. However, EGM signals patterns associated with rotors are still unknown, and the tools for characterizing fractionated EGM signals are still under development.

This paper presents a method for automatically classifying EGM signals into four classes by a combination of relevant features. The signals are assigned one of the four classes, in accordance with the system proposed by Schilling et al. [15]. The goal of this approach is to develop software for automatically identifying the class of each signal; then, use this information; and, finally, map the spatial distribution of the classes. We test the method using 2D and 3D simulations of fibrillatory conduction to associate the spatial distribution of the classes with the rotor's vortex – considered a critical conduction zone.

2. Materials and methods

This study was divided into two stages. In the first one, real EGM signals were collected and then several methods were employed to extract the features – e.g. time domain, frequency domain, time-frequency, nonlinear-based technique and information theory features. Subsequently, the relevant subset of features showing the best performance for classification was selected. The selected subset of features was processed by a *K*-nearest neighbors (K-NN) classifier. Afterwards, the results were cross-validated. During the second stage, the classifier was used to evaluate levels of EGM fractionation in two models simulating AF episodes. The goal of this method is to find the relation between levels of EGM fractionation and the spatial distribution of arrhythmogenic substrate in the model.

2.1. Experimental dataset of EGM

The EGM dataset is composed of 429 EGM records acquired from 11 AF patients. Among them, there were eight cases of persistent AF and three cases of paroxysmal AF. All signals were collected during AF, in PVI procedures. Such dataset was previously provided by the Staedtisches Klinikum Karlsruhe in Germany [20]. They used several catheters: a Lasso multipolar circular catheter (10 polar, Biosense Webster, Diamond Bar, USA), an Optima (14 polar, St. Jude Medical, St. Paul, USA) and an Orbiter (14 polar, Bard Electrophysiology, Lowell, USA). EGM signals were recorded with a sampling rate of 1.2 kHz. All EGM signals were digitally down-sampled to 1 kHz and were filtered with a band pass filter at [0.3, 250] Hz, and they were preprocessed by a wavelet decomposition filter [21] in order to remove baseline wandering.

Two experienced electrophysiologists, working at different centers, independently annotated the database. They classified the signals according to the preset fractionation; the non-fractionated EGM signals were considered to be Class 0 (C0). The fractionated signals were divided into three classes of fractionation: mild, intermediate and high; C1, C2, and C3, respectively. In addition, after a visual inspection by experts, signals having the following characteristics were also sorted out: (i) low quality and/or very low voltage signals, (ii) signals superimposed on ventricular far-field components, (iii) signals that remain non-stationary over the whole five-second recording period. The resulting 429 EGM signals were selected from 2824 signals after this discarding process. Only signals that had been assigned to the same class by both physicians were used to create the database [15]. Those 429 signals were distributed into fractionation classes: 153 signals in C0, 75 signals in C1, 148 signals in C2, and 53 signals in C3 (see Fig. 1).

2.2. Experimental design

Fig. 2 shows a diagram of the steps involved in this work. (i) 92 features were extracted by using various methods (see below). (ii) The genetic algorithms found an optimal subset of features. Feature selection was implemented in a wrapper scheme with the classifier. (iii) The K-NN classifier was used to detect fractionated EGM signals in both, the 3D and the 2D, atrial models using the given labeled EGM database as training set. For the K-NN classifier, $K = 3$ was heuristically selected. (iv) An electroanatomical map was constructed from the results of the classification to establish the relation between levels of EGM fractionation and the location of the rotor vortex in the models.

2.3. EGM signal characterization feature extraction methods

A feature extraction routine was carried out to characterize the signal and later use these features in a classifier. The extracted features were divided into 5 groups depending on their origin. Below is a brief explanation of each group. Further annotation can be found in the supplementary material.

2.3.1. Group 1: Features based on temporal information and detecting local activation waves

The first group of features is based on the characterization of near- and far-field potentials in the EGM. Here we aim to search for electrical activity in the signal. According to Houben et al. [22], a local activation wave (*LAW*) is mainly composed of near field potentials, and it is detected by using a method based on wavelet transform. The *LAWs* were detected as follow: (i) The continuous wavelet transform (CWT) was performed. The second derivative of the Gaussian function was used as a mother wavelet. (ii) The reconstruction of scales 1, 2 and 3 was computed. (iii) A moving window integrator filter was implemented to join adjacent inter-segments if the inter-segment is shorter than the window time (40 ms, heuristically selected). (iv) We applied an adaptive threshold, based on the algorithm described by Pan and Tompkins [23]. (v) Signal segments with amplitude greater than the threshold were detected. Their maximum values in each segment were marked as *centers*. Features such as zero crossings (Zc), maxima, minima, mean and standard deviation of the distance between *LAW* centers were obtained (see Fig. 3).

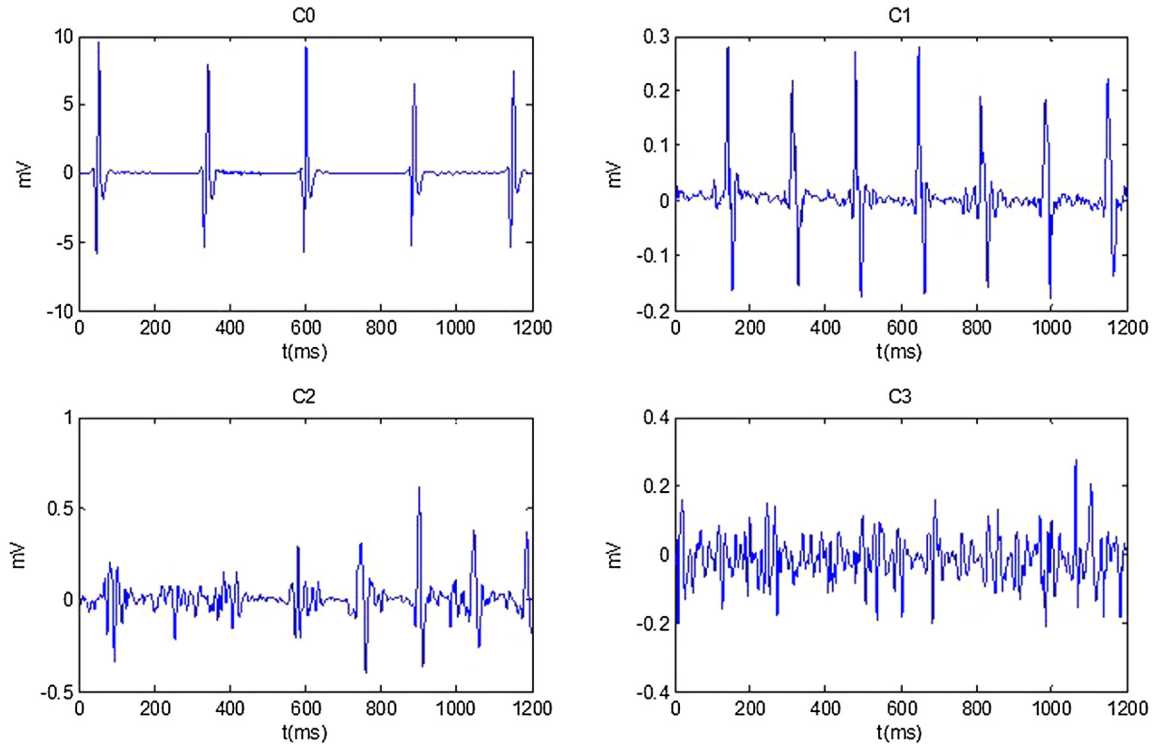

Fig. 1. Four samples of atrial EGMs. They represent the four levels of fractionation: C0, organized atrial activity; C1, mild fractionation; C2, intermediate fractionation; and C3, high degree of fractionation.

Fig. 2. Methodological steps in the process.

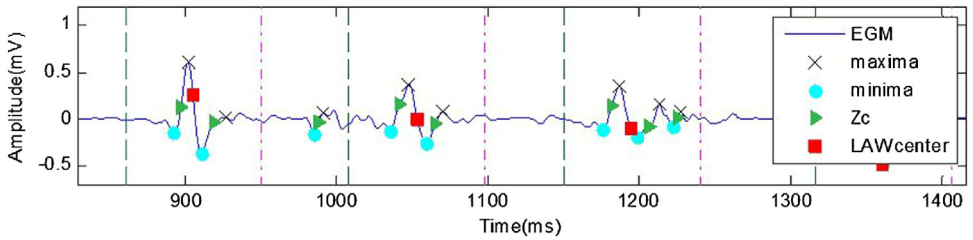

Fig. 3. *LAW* Feature extraction of Local Activation Waves. The green dashed line represents the beginning of the *LAW*; the magenta dash-dot line marks the end of the *LAW*. The figure shows the center, zero crossings (Zc), maxima and minima of each *LAW*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.3.2. Group 2: Features based on the LAW morphology

The morphological information of the *LAW* was extracted by processing the data within the *LAW* windows. Two different actions were performed to this end:

- The similarity index (SI) was calculated on the basis of the index proposed by Faes et al. [24] with some modifications to the synchronization method. The correlation and the cosine distance between *LAW* segments were used to calculate two different SI indexes.
- A template was created through the sum of all *LAWs*. Subsequently, it was the source to extract descriptive features such as amplitudes, peaks and zero crossing (Zc), mean, standard deviation, maximum and minimum, energy, and singular value decomposition points. For further information, see supplementary material.

In order to detect the time gap between a pair of waves, we used cross-correlation. Afterwards, the *LAWs* were synchronized to calculate the SI and the template. Such process was based on the location of the maximum point detected in the cross-correlation signal (α). All synchronized segments were normalized (see Fig. 4).

2.3.3. Group 3: Features based on segments with continuous local electrical activity

Our approach was based on the work by Kremen et al. [25], We compute all the Segments with Continuous Local Electrical Activity (*SLEA*). First, we used CWT and reconstruct the signal using the scales 1, 2 and 3 (Fig. 5a and b). Second, primary local activ-

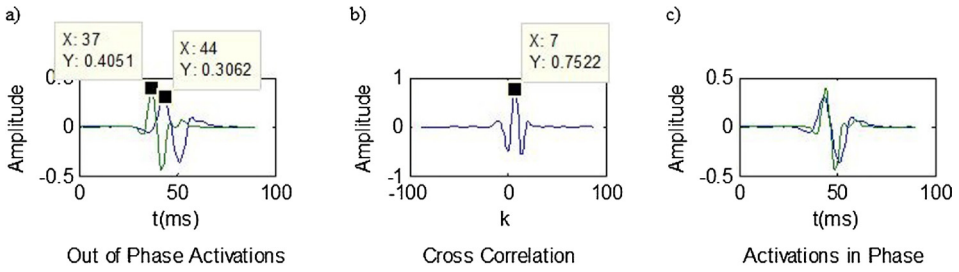

Fig. 4. Synchronization procedure. (a) Two unsynchronized *LAW*. (b) Cross correlation signal obtain from two waves. Based on this signal, we can calculate the phase shift (α) between the two waves – for this example $\alpha = 7$. (c) In-phase waves after applying a time correction.

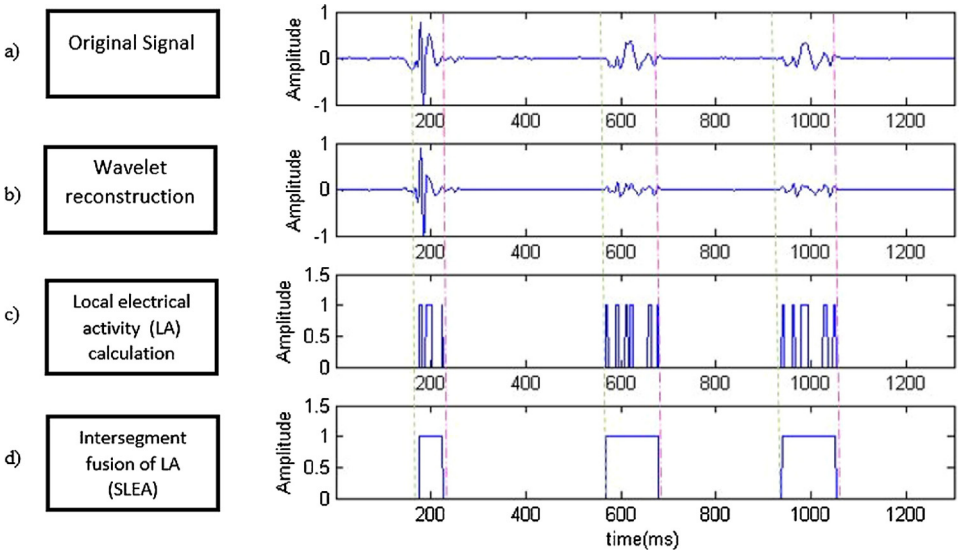

Fig. 5. Sequence of the search for segments with local electrical activity (*SLEA*) to measure the width of the activations. The green dashed line represents the beginning of the *SLEA*; the magenta dash-dot line marks the end of the *SLEA*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ity segments (*PLAS*) are detected in the reconstructed signal using a threshold (Fig. 5c). SLEAs were defined as adjacent *PLAS* whose inter-segments space is less than 40 ms (Fig. 5d). Two measures were computed: t_{SLEA} was defined as the total time of the sum of all *SLEAs*; and t_{LA}, the time of the sum of all *PLAS*. The algorithm extracted features based on *SLEAs*, t_{SLEA} and the relation between t_{SLEA} and t_{LA}.

2.3.4. Group 4: Features based on the frequency domain and the on time-frequency domain

The periodogram that shows the power spectral density of the signal was computed with a rectangular window and a number of points equal to the next power of two greater than the signal length. Subsequently, such periodogram was used to extract features – e.g. center frequency, mean frequency, maximum frequency, quartiles in the frequency spectrum, skewness, and total energy spectrum. Besides, we implemented the algorithm described by Everett et al. [26] to compute the dominant frequency (*DF*) and the regularity index (*RI*). The latter were calculated from an signal envelope created by following the procedure proposed by Botteron and Smith [27]. We also applied the discrete wavelet transform of the signal at 8 scales and Daubechies 6 mother wavelet. The energy of each of the detail coefficients and the approximation coefficient were obtained as features.

2.3.5. Group 5: Features based on information theory and non-linear dynamics

Several methods based on information theory were used – e.g. autocorrelation, singular value decomposition (*SVD*) and mutual information (*MI*) between two subsegments of 300 ms extracted from the signal. Additionally, approximate entropy (*ApEn*) was calculated [28]. The latter is a nonlinear statistical technique used to quantify the complexity in a signal. *ApEn* was calculated in accordance with the parameters suggested by Ugarte et al. [29], $m = 3$ and $r = 0.38$.

2.4. Feature selection

The feature selection was carried out using genetic algorithms to find an optimal subset of features providing the best performance. A genetic algorithm is a stochastic global search method that is used to solve complex problems [30]. The algorithm creates a *population* of possible solutions to the problem that enables them to *evolve* for several generations to find an optimal solution. We used the MATLAB Genetic Algorithm toolbox and set the following parameters: PopulationSize: 20, Generations: 600, TimeLimit: 3600, FitnessLimit: 0.05, StallGenLimit: 50, TolFun: 0.01. We used the error rate of K-NN as the cost function, and applied cross validation with 10 folds.

2.5. Classification of database signals

In order to classify the database signals, we implemented the K-NN method [31]. K-NN is a supervised classification method used in statistical estimation and pattern recognition. It is a nonparametric technique that estimates the probability that an element x belongs to a class C_i based on training data. A sample is classified by a majority vote of its neighbors – i.e. the sample is assigned the most common class among the K-nearest neighbors measured by a function of distance. The distance function we used was Euclidean distance, and the number of neighbors was $K = 3$. This parameter was heuristically selected. In order to avoid overtraining, n-cross validation was performed. Finally, randomly generated partitions were selected from the whole dataset and $n = 10$ subsets were created.

2.6. Rotor vortex identification in a 3D atrial model and in a 2D atrial model

A realistic model of human atria [32] was used, and an AF episode was generated by the S1–S2 protocol. First, a train of stimuli with a basic cycle length of 1000 ms was applied for a period of 5 s in the area to trigger the sinus rhythm (S1). Second, after the last beat of the sinus stimulus, a burst of 6 ectopic beats (S2) at high frequency was delivered into the right superior pulmonary vein. We simulated 5 s of an AF episode. Finally, we simulated unipolar EGM at different points of the atrial surface under conditions of uniform intracellular anisotropic resistivity, as described in Tobon et al. [32].

In previous work there have been detailed descriptions of 3D human atria models, the method for simulating an AF episode (maintained by a rotor in the posterior wall of the left atrium), and atrial surface EGM reconstruction [33]. By altering the ionic properties, EGM fractionation can also be obtained in simulated from models reproducing AF [34]. A detailed description of the relation between simulated EGMs and arrhythmic activation patterns has previously been provided [33,35].

In our case, during the simulated AF episode, a rotor in the posterior wall of the left atrium was observed. This activity caused the appearance of electrotonic and irregular action potentials. Likewise, fractionated atrial EGM were found in areas near the rotor's vortex. Such EMGs showed potentials composed of two or more deflections. The rotor wavefront surrounded the pivot point without depolarizing it completely. This results in multiple low amplitude deflections in the EGM. Then, we identified the area where the rotor turns in the posterior wall of the left atrium. EGM signals obtained from this area were used in this study (see Fig. 6a). We perform a conversion of the the selected region of the model from the 3D to the 2D coordinate system (x, z). A very low dispersion in y benefit this process. Using the 2D area, we mapped the classification of the EGMs – converted to bipolar EGM – of the simulated model.

On the other hand, we have also developed a 2D model of human atrial tissue. The tissue surface was discretized into a 150×150 hexahedral mesh (22,500 elements and 45,602 nodes). The spatial resolution was 0.4 mm. The electrophysiological model was integrated into the 2D virtual model. Additionally, the tissue was considered to be isotropic. A conductivity of 0.3 S/cm was set in order to obtain a realistic conduction velocity of 60 cm/s. A rotor was generated by S1–S2 cross-field stimulation protocol. The stimuli pulses were rectangular (2×3 cm). The S1 stimulus was applied at the left boundary of the model (2 ms of duration, 6 mA of amplitude). The S2 stimulus was triggered 40 ms after S1 on a corner of the model. After following the protocol, a four-second stable rotor was observed in the 2D model (see Fig. 7a).

The pseudo EGM generated by both models originally were unipolar. To compare them with the signals from the real EGM database, the pseudo EGM needed to be converted to bipolar signals. This task was completed by creating a mesh to simulated bipolar electrodes with 1 mm spacing. The difference between two adjacent signals in the mesh was thus calculated, and the bipolar signals were obtained.

Table 1
Performance of classification by genetic algorithms with K-NN at four EGM classes. Correctly classified EGM among all groups (Cr); Sensitivity (Se); and Specificity (Sp). Se and Sp were calculated for each group.

Class	Cr (%)	Se (%)	Sp (%)
C0		92.83 ± 6.38	98.57 ± 1.84
C1	92.08 ± 5.35	90.71 ± 8.74	97.21 ± 3.48
C2		95.29 ± 4.53	94.65 ± 5.08
C3		83.33 ± 17.64	98.68 ± 1.87

3. Results

A total of 92 features were extracted from 429 real EGM signals in the database. They formed a matrix $\Theta \in \mathbb{R}^{429\times92}$. Such features are listed in the supplementary material.

3.1. Feature selection

The optimal subset of 92 features was found by genetic algorithms. Thus, the number of features used by the classifier was reduced to a new matrix $\Theta \in \mathbb{R}^{429\times42}$. Relevant features that enable to characterize the signal – e.g. the similarity index, approximate entropy, singular value decomposition of the templates, and the standard deviation of the autocorrelation signal were extracted. The features listed in the optimal subset can be found in the supplementary material.

3.2. Classification of real EGM signals from the database

An analytical assessment of the classification was performed by calculating the percentage of correctly samples classified (Cr). Besides, the sensitivity (Se) and specificity (Sp) were determined for each one of the fractionation levels (C0, C1, C2 and C3). The results are shown in Table 1.

3.3. Classification of virtual EGM from simulated AF episode

We developed a piece of software that simulated an AF episode in the 3D model and 2D models to illustrate a possible application of automatically classifying EGM according to their level of fractionation.

The EGM bipolar signals of the 3D model and the 2D model were tagged using a K-NN classifier. The training set consists of the signals from the database labeled by experts. The results of automatic classification were displayed on color maps.

The output of the classification of EGM signals from the 3D model and 2D model show that the rotor vortex area is associated with class 3 signals (Figs. 6 and 7). These EGMs present disorganized electrical activity. These results suggest that fractionation levels can be associated with the rotor. Signals with organized and non-fragmented activity (single potentials = Class 0) are related to a regular wave front in the model. Likewise, signals with continuous and fragmented potentials (Class 3) are linked to pivot points. Class 3 signals from the models show fragmented potential with several direction changes and high frequency respect with other classes. Therefore, these findings indicate that there is a similarity between the morphology of the virtual and real EGMs.

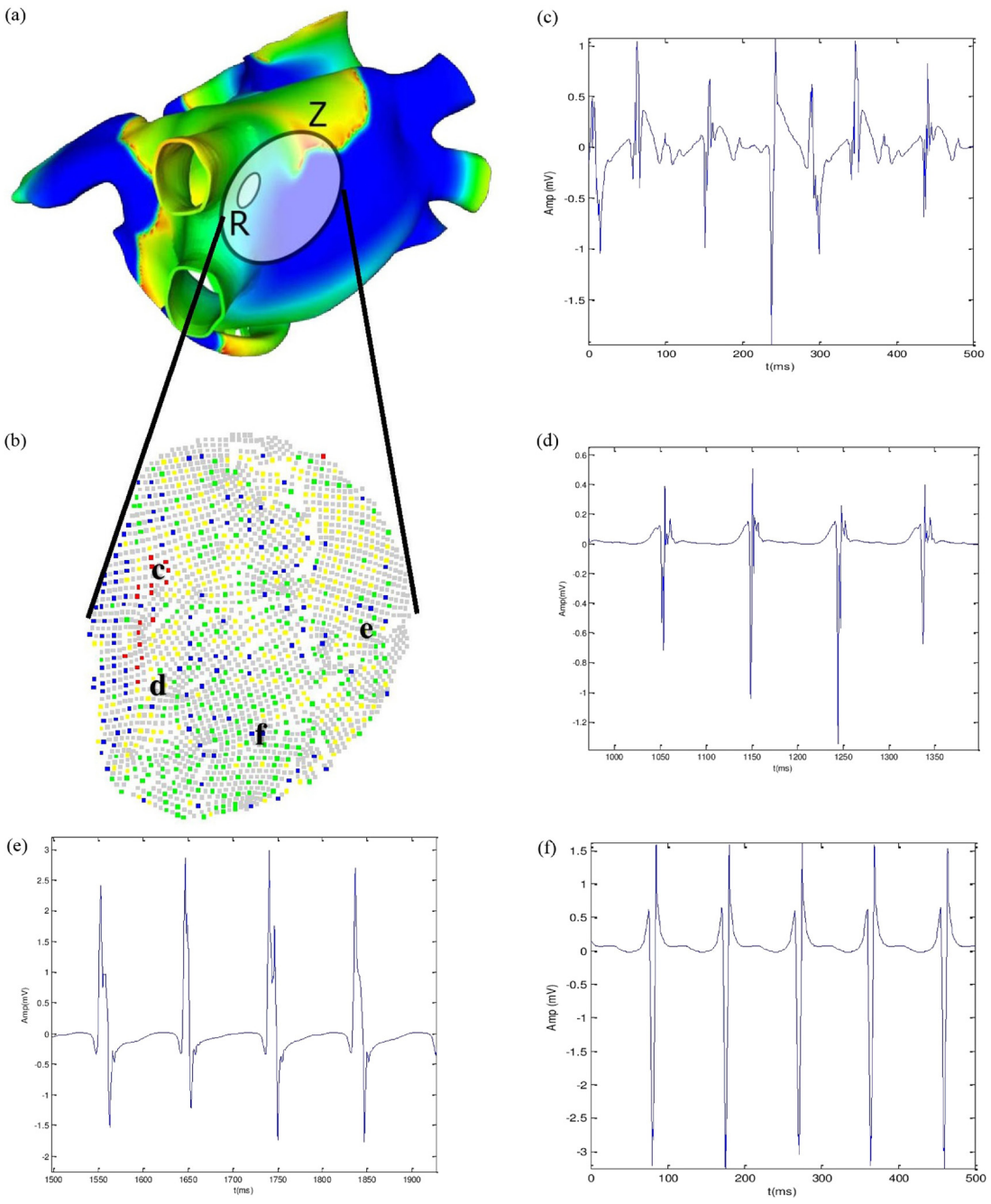

Fig. 6. Classification of signals from the posterior wall of the left atrium of the 3D model. EGM classification is represent in a color map: blue (C0), green (C1), yellow (C2), and red (C3). Pixels in gray are not associated with any bipolar signal. (a) Action potential propagation map on the posterior wall of the left atrium view. *Z* denotes the scanned area ((*x*, *z*) projection), and *R*, the rotor's vortex. (b) Color map obtained with the K-NN classifier. Importantly, the area where the signals are highly fragmented corresponds to the area of the rotor's vortex in the activation map. (c) Sample of C3 Signals. It can be see that the signal presents high level of fractionation. (d) Sample of C2 signal. (e) Sample of C1. (f) C0 signal. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

PVI is the first line treatment for paroxysmal AF. Also, PVI in combination with CFAE constitute another strategy against persistent AF. However, the efficacy of targeting CFAE sites is debated, mainly because there is considerable uncertain with regards to the meaning of different CFAE morphologies. CFAEs are believed to be related to sites that present slow conduction, structural obstacles, wave collisions, or pivot points. Our findings suggest that supervised classifier trained on real EGM signals is able to detect different EGM morphologies and locate critical sites in AF.

In our study we have described an automated implementation of a supervised learning approach to classify real EGM signals recorded during radiofrequency ablation of AF. Such classification divided the signals into four classes according to their levels of fractionation. We have also probed into the relation between EGM fractionation and a rotor's vortex during simulated episodes of chronic AF. Our work presents several results: (i) An evaluation and selection of a subset of the optimum features for classifying real EGM into 4 levels of fractionation using a K-NN classifier. (ii) The use of real EGM signals labeled by experts to train the classifier. The trained classifier was used in a software application to detect

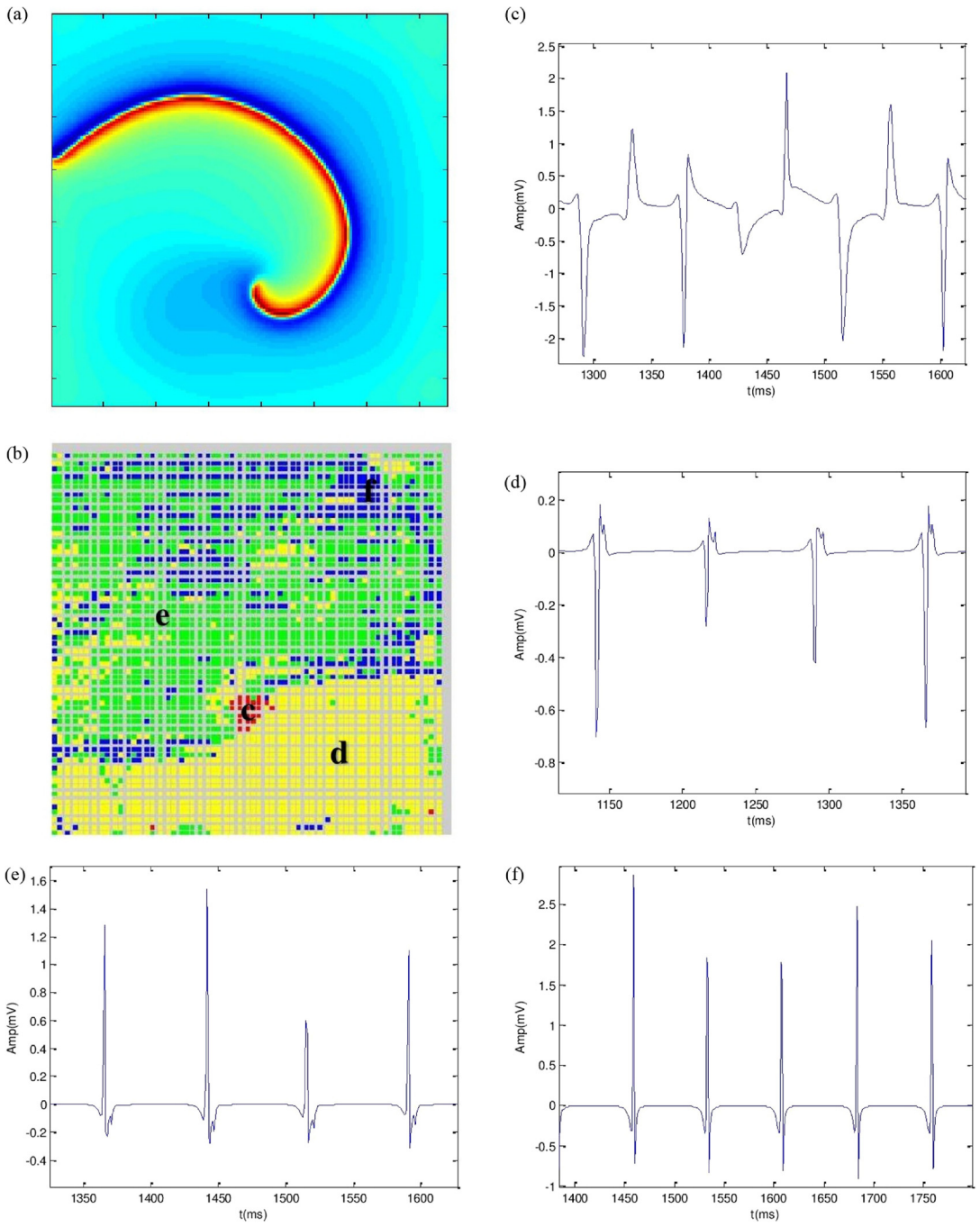

Fig. 7. Classification of signals in the 2D model. (a) Action potential propagation map of a 2D model of atrial tissue. (b) Color map obtained with the K-NN classifier (C0 = blue, C1 = green, C2 = yellow and C3 = red). It is important to note that the area marks in red corresponds to the rotor vortex. (c), (d), (e) and (f) are samples of signal classified as C3, C2, C1 and C0, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the location of the rotor's vortex in a 3D and a 2D models of atrial tissue during simulated episodes of chronic AF.

In commercial systems, features extraction from EGMs, CFAE detection is generally carried out on the basis of time-domain morphology analysis [36]. However, commercial systems do not distinguish different morphologies. They are limited to detecting the number or the deflection in a time window according to pre-selected parameters and distinguish between fragmented and non-fragmented signals [10]. Therefore, the levels of fractionation found in EGM signals are not well defined. Several authors have shown that a classification based on different levels of fractionation could be a better approach in a study of critical sites for AF [37,15]. Besides, non-linear features [17,29,38] could improve the detection of critical sites associated with rotors. A combination of EGM features is a promising way to discriminate arrhythmogenic substrates [39]. Our method takes a combination of 42 features as the input for a K-NN classifier, including linear and non-linear features, to achieve better performance in EGM classification.

Detecting highly fragmented EGMs is a key factor in the identification of critical sites that could be targeted for AF ablation. Faes et al. [24] created a classification of 3 levels of fractionation using a similarity index. They achieved an accuracy value of 85.5%. Kremen et al. [25] implemented an algorithm based on SLEAs to sort fractionated electrograms (C3 in our case). The results were

82% for sensitivity and 90% for specificity within the highest level of fractionation. Hunter et al. [40] evaluated the accuracy of the algorithms of commercial systems with user-defined parameters. They divided the signals into five levels of fractionation to detect the highest level. Their results were 91% of sensitivity and 91% of specificity. The outcomes from the above mentioned studies were obtained using a single feature extraction method: Faes et al. [24] the regularity index; Kremen et al. [25] used measures based on segments with local electrical activity; and Hunter et al. [40] evaluated the fractionation under amplitude and slope criteria in the time domain.

On the other hand, Schilling et al. [15] implemented a fuzzy decision tree to classify fractionated EGMs using the same dataset that was used in this work plus a combination of different features. They obtained a correct rate of $80.65 \pm 3.3\%$. In our work, the performance of the classifier provided a correct rate of $92.08 \pm 5.35\%$, a sensitivity of $90.54 \pm 9.32\%$, and a specificity of $97.28 \pm 3.07\%$. This performance hinges on the fact that the classifier works with heterogeneous information regarding the EGM fractionation/regularity. Such data results from the extraction and selection of several independent features of each EGM signal. Our results indicate that the selected features are suitable for differentiating levels of fractionation, even with a simple and fast algorithm such as the K-NN method.

The 2D and 3D atrial models were used to illustrate a possible application of the proposed procedure. In-silico studies enable a reproducible evaluation of AF conduction (with high resolution) for detecting action potential propagation. For now, this is not feasible in humans. Our results with simulated data are consistent with the EGM morphologies described by Konings et al. [19]. They made a connection between fragmented potential at pivot points and single potential at sites with rapid uniform conduction. In our study, the EGM signals over the rotor area, identified by the action potential activation maps, were classified in the group with the highest degree of fractionation (see Fig. 6c). These results were obtained by applying the K-NN classifier to signals recorded from the 3D and 2D model. Real EGM signals were used in the training set. Furthermore, our classifier results show that signals from simulated electrodes away the rotor's vortex, exhibit patterns that correspond to the class C0 in the real dataset (non-fragmented EGM – Figs. 6f and 7 f). In addition, Fig. 6c and 7 c show that the signals identified as C3 by the classifier in our models are associated with the position of the rotor's vortex and they are not present in other locations. We therefore suggest that the detection of such signals with appropriate levels of sensitivity and specificity could be used for identifying critical sites for AF. The models in this study have been validated only in one simulated data set for each model, and further confirmation is necessary. In future studies, several simulations will be performed.

One of the methods for locating critical sites for AF is to track activation patterns [41]. Nevertheless, in real clinical practice it is difficult to use activation maps. This is due to the low spatial resolution, the effect of highly disorganized signals, and the large amount of computation time required for evaluating the map for each activation wave. By contrast, our results suggest that the proposed classification method based on extracting and selecting various types of EGM features can serve as the basis for quick automatic detection of critical sites. The computational time required to calculate the 42 selected features is 1.268 s for one 5000 ms signal and 162.96 s for 429 signals. The time was measured using MatLab 2013b in a desktop PC running (64 bits) Windows 8 on a Intel Core I7 processor and 8 GB of RAM.

The goal of our study has been to classify the signals into four fractionation levels in accordance with previously used criteria (Retrospective study). We then used a dataset reported in the literature and the results obtained by their classification methods [15]. The limitations of our study are related to the database that was used. The visual fractionation assessment employed to label the signals introduces subjectivity into the method, as each specialist has his own perception of fractionation. Additionally, we are not sure whether visual identification of fractionation accurately identifies true atrial substrate. On this point, it is important to take into account that levels of fractionation and EGM morphologies are often only vaguely described, or differently defined in the literature. This makes very difficult to tell fractionation levels apart, even for electrophysiologists. Another limitation of our study is related to the use of only four EGM classes. Besides, some signals were discarded by visual inspection when selecting the dataset. Subsequently, all signals in the training set were temporally stable (had the same level of fractionation over a period of 5 s). This implies that some signals will be not suitable for classification during clinical application of this approach. Future studies will aim to develop a method for discarding such unsuitable signals. Semi-supervised learning could be a feasible approach to reduce dependence on the training set [42].

The relevance of locating EGM patterns is supported by several studies that have indicated that some particular fractionated morphologies are likely to represent drivers of AF [43,19]. Moreover, experimentation on isolated animal hearts has shown that areas with the most fractionated EGM signals coexist on the periphery of the most rapid and less fractionated places [44,34]. This may lead to the localization of AF sources. It also implies that the identification of different classes across the atrial surface can become an adequate diagnostic support tool to track critical AF sites. Future studies should address how the patterns are related to AF substrate and how AF terminates after ablating an atrial region depending on a particular EGM class.

5. Conclusion

Our work has indicated that highly fractionated EGM signals can be detected using a feature selection system and a K-NN classifier trained on a set of real EGM tagged into four fractionation levels. Our method enables different EGM morphologies to be located over the atrial surface. This could be a tool for developing systems that guide ablation procedures. In order to test the strength of this method for locating critical sites in a real context, we need a study comprising a complete dataset of EGMs from different patients and the signals from the whole atrial surface of the same individuals.

Acknowledgments

We would like to acknowledge the expert advice and the data provided by Prof. Olaf Dossel and his laboratory team, from the Institute of Biomedical Engineering at Karlsruhe Institute of Technology in Germany, and by Dr. Armin Luik, from the Cardiology and Internal Medicine Departments at Karlsruhe City Hospital in Germany.

This work was supported by COLCIENCIAS (Departamento Administrativo de Ciencia, Tecnología e Innovación de la República de Colombia) though grants No. 121056933647, No. 121065741044, *Jovenes Investigadores* and *Doctoral program*. It was also aided by Czech Science Foundation (GACR), postdoctoral research project GACR P103/11/P106. Institutional resources for research were supported by the Czech Technical University in Prague, Czech Republic and Universidad Pontificia Bolivariana and Instituto Tecnológico Metropolitano in Medellín, Colombia.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bspc.2017.06.005.

References

[1] E.J. Benjamin, P.A. Wolf, R.B. Agostino, H. Silbershatz, W.B. Kannel, D. Levy, Impact of atrial fibrillation on the risk of death the Framingham heart study, Circulation 98 (10) (1998) 946–952.
[2] J.J. Goldberger, J. Ng, Practical Signal and Image Processing in Clinical Cardiology, Springer, 2010.
[3] A.J. Camm, P. Kirchhof, G.Y. Lip, U. Schotten, I. Savelieva, S. Ernst, I.C. Van Gelder, N. Al-Attar, G. Hindricks, B. Prendergast, et al., Guidelines for the management of atrial fibrillation, Eur. Heart J. 31 (2010) 2369–2429.
[4] V. Barbaro, P. Bartolini, G. Calcagnini, F. Censi, A. Michelucci, S. Morelli, Mapping the organization of human atrial fibrillation using a basket catheter, Computers in Cardiology, 1999, IEEE (1999) 475–478.
[5] J. Ng, A.H. Kadish, J.J. Goldberger, Technical considerations for dominant frequency analysis, J. Cardiovasc. Electrophysiol. 18 (7) (2007) 757–764.
[6] K. Nademanee, J. McKenzie, E. Kosar, M. Schwab, B. Sunsaneewitayakul, T. Vasavakul, C. Khunnawat, T. Ngarmukos, A new approach for catheter ablation of atrial fibrillation: mapping of the electrophysiologic substrate, J. Am. Coll. Cardiol. 43 (11) (2004) 2044–2053.
[7] J. Seitz, J. Horvilleur, J. Lacotte, Y. Mouhoub, F. Salerno, A. Moynagh, O. Darach, M. Monchi, L. Curel, A. Pisapia, Automated detection of complex fractionated atrial electrograms in substrate-based atrial fibrillation ablation: better discrimination with a new setting of carto algorithm, J. Atrial Fibrill. 6 (2) (2013).
[8] H. Oral, A. Chugh, E. Good, A. Wimmer, S. Dey, N. Gadeela, S. Sankaran, T. Crawford, J.F. Sarrazin, M. Kuhne, N. Chalfoun, D. Wells, M. Frederick, J. Fortino, S. Benloucif-Moore, K. Jongnarangsin, F. Pelosi, F. Bogun, F. Morady, Radiofrequency catheter ablation of chronic atrial fibrillation guided by complex electrograms, Circulation 115 (20) (2007) 2606–2612.
[9] J. Chen, Y. Lin, L. Chen, J. Yu, Z. Du, S. Li, Z. Yang, C. Zeng, X. Lai, Q. Lu, B. Tian, J. Zhou, J. Xu, A. Zhang, Z. Li, A decade of complex fractionated electrograms catheter-based ablation for atrial fibrillation: literature analysis, meta-analysis and systematic review, IJC Heart Vessels (2014) 1–10.
[10] D.H. Lau, B. Maesen, S. Zeemering, S. Verheule, H.J. Crijns, U. Schotten, Stability of complex fractionated atrial electrograms: a systematic review, J. Cardiovasc. Electrophysiol. 23 (9) (2012) 980–987.
[11] D.H. Lau, B. Maesen, S. Zeemering, P. Kuklik, A. van Hunnik, T.A. Lankveld, E. Bidar, S. Verheule, J. Nijs, J. Maessen, et al., Indices of bipolar complex fractionated atrial electrograms correlate poorly with each other and atrial fibrillation substrate complexity, Heart Rhythm 12 (7) (2015) 1415–1423.
[12] M. Omer Berenfeld, J. Jalife, Complex fractionated atrail electrograms is this the beast to tame in AF, NIH Public Access 4 (4) (2011) 426–428.
[13] T.P. Almeida, J.L. Salinet Jr., G.S. Chu, G.A. Ng, F.S. Schlindwein, C. Sciences, G. Hospital, Different definitions of complex fractionated atrial electrograms do not concur with the clinical perspective, Comput. Cardiol. 40 (2013) 1055–1058.
[14] J.L. Wells, R.B. Karp, N.T. Kouchoukos, W.A. Maclean, T.N. James, A.L. Waldo, Characterization of atrial fibrillation in man: studies following open heart surgery, Pacing Clin. Electrophysiol. 1 (4) (1978) 426–438.
[15] C. Schilling, M. Keller, D. Scherr, T. Oesterlein, M. Haïssaguerre, C. Schmitt, O. Dössel, A. Luik, Fuzzy decision tree to classify complex fractionated atrial electrograms, Biomed. Eng./Biomedizinische Technik (2015).
[16] J. Jalife, O. Berenfeld, M. Mansour, Mother rotors and fibrillatory conduction: a mechanism of atrial fibrillation, Cardiovasc. Res. 54 (2) (2002) 204–216.
[17] A.N. Ganesan, P. Kuklik, D.H. Lau, A.G. Brooks, M. Baumert, W.W. Lim, S. Thanigaimani, S. Nayyar, R. Mahajan, J.M. Kalman, et al., Bipolar electrogram shannon entropy at sites of rotational activation implications for ablation of atrial fibrillation, Circulation: Arrhythm. Electrophysiol. 6 (1) (2013) 48–57.
[18] A. Orozco-Duque, J.P. Ugarte, C. Tobón, J. Saiz, J. Bustamante, Approximate entropy can localize rotors, but not ectopic foci during chronic atrial fibrillation: a simulation study, Computing in Cardiology Conference (CinC), 2013, IEEE (2013) 903–906.
[19] K.T. Konings, J.L. Smeets, O.C. Penn, H.J. Wellens, M.A. Allessie, Configuration of unipolar atrial electrograms during electrically induced atrial fibrillation in humans, Circulation 95 (5) (1997) 1231–1241.
[20] C. Schilling, Analysis of Atrial Electrograms, vol. 17, KIT Scientific Publishing, 2012.
[21] D. Zhang, Wavelet approach for ECG baseline wander correction and noise reduction, IEEE-EMBS 2005. 27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005, IEEE (2005) 1212–1215.
[22] R.P. Houben, N. De Groot, M. Allessie, et al., Analysis of fractionated atrial fibrillation electrograms by wavelet decomposition, IEEE Trans. Biomed. Eng. 57 (6) (2010) 1388–1398.
[23] J. Pan, W.J. Tompkins, A real-time QRS detection algorithm, IEEE Trans. Biomed. Eng. (3) (1985) 230–236.
[24] L. Faes, G. Nollo, R. Antolini, F. Gaita, F. Ravelli, A method for quantifying atrial fibrillation organization based on wave-morphology similarity, IEEE Trans. Biomed. Eng. 49 (12) (2002) 1504–1513.
[25] V. Křemen, L. Lhotská, M. Macaš, R. Čihák, V. Vančura, J. Kautzner, D. Wichterle, A new approach to automated assessment of fractionation of endocardial electrograms during atrial fibrillation, Physiol. Meas. 29 (12) (2008) 1371.
[26] T.H. Everett IV, L.-C. Kok, R.H. Vaughn, J.R. Moorman, D.E. Haines, Frequency domain algorithm for quantifying atrial fibrillation organization to increase defibrillation efficacy, IEEE Trans. Biomed. Eng. 48 (9) (2001) 969–978.
[27] G.W. Botteron, J.M. Smith, A technique for measurement of the extent of spatial organization of atrial activation during atrial fibrillation in the intact human heart, IEEE Trans. Biomed. Eng. 42 (6) (1995) 579–586.
[28] S.M. Pincus, Approximate entropy as a measure of system complexity, Proc. Natl. Acad. Sci. U. S. A. 88 (6) (1991) 2297–2301.
[29] J.P. Ugarte, A. Orozco-Duque, C. Tobón, V. Kremen, D. Novak, J. Saiz, T. Oesterlein, C. Schmitt, A. Luik, J. Bustamante, Dynamic approximate entropy electroanatomic maps detect rotors in a simulated atrial fibrillation model, PLOS ONE 9 (12) (2014) e114577.
[30] C.R. Houck, J. Joines, M.G. Kay, A genetic algorithm for function optimization: a matlab implementation, NCSU-IE TR 95 (09) 1995.
[31] T.M. Cover, P.E. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theory 13 (1) (1967) 21–27.
[32] C. Tobón, J.F. Rodríguez, J.M. Ferrero, F. Hornero, J. Saiz, Dominant frequency and organization index maps in a realistic three-dimensional computational model of atrial fibrillation, Europace 14 (Suppl. 5) (2012) v25–v32.
[33] C. Tobón, C.A. Ruiz-Villa, E. Heidenreich, L. Romero, F. Hornero, J. Saiz, A three-dimensional human atrial model with fiber orientation, electrograms and arrhythmic activation patterns relationship, PLOS ONE 8 (2) (2013) e50883.
[34] S. Zlochiver, M. Yamazaki, J. Kalifa, O. Berenfeld, Rotor meandering contributes to irregularity in electrograms during atrial fibrillation, Heart Rhythm 5 (6) (2008) 846–854.
[35] J. Ugarte, C. Tobón, A. Orozco-Duque, M. Becerra, J. Bustamante, Effect of the electrograms density in detecting and ablating the tip of the rotor during chronic atrial fibrillation: an in silico study, Europace (2015).
[36] D. Scherr, D. Dalal, A. Cheema, A. Cheng, C.A. Henrikson, D. Spragg, J.E. Marine, R.D. Berger, H. Calkins, J. Dong, Automated detection and characterization of complex fractionated atrial electrograms in human left atrium during atrial fibrillation, Heart Rhythm 4 (8) (2007) 1013–1020.
[37] G. Nollo, M. Marconcini, L. Faes, F. Bovolo, F. Ravelli, L. Bruzzone, An automatic system for the analysis and classification of human atrial fibrillation patterns from intracardiac electrograms, IEEE Trans. Biomed. Eng. 55 (9) (2008) 2275–2285.
[38] A. Orozco-Duque, D. Novak, V. Kremen, J. Bustamante, Multifractal analysis for grading complex fractionated electrograms in atrial fibrillation, Physiol. Meas. 36 (2015) 2269–2284.
[39] F. Ravelli, M. Masè, Computational mapping in atrial fibrillation: how the integration of signal-derived maps may guide the localization of critical sources, Europace 16 (5) (2014) 714–723.
[40] R.J. Hunter, I. Diab, G. Thomas, E. Duncan, D. Abrams, M. Dhinoja, S. Sporton, M.J. Earley, R.J. Schilling, Validation of a classification system to grade fractionation in atrial fibrillation and correlation with automated detection systems, Europace 11 (12) (2009) 1587–1596.
[41] S.M. Narayan, M. Wright, N. Derval, A. Jadidi, A. Forclaz, I. Nault, S. Miyazaki, F. Sacher, P. Bordachar, J. Clémenty, P. Jaïs, M. Haïssaguerre, M. Hocini, M. Sanjiv, M. Narayan, M. Wright, N. Derval, A. Jadidi, A. Forclaz, I. Nault, S. Miyazaki, F. Sacher, P. Bordachar, J. Clémenty, P. Jaïs, M. Michel Haïssaguerre, M. Hocini, Classifying fractionated electrograms in human atrial fibrillation using monophasic action potentials and activation mapping: evidence for localized drivers, rate acceleration, and nonlocal signal etiologies, NIH Public Access 8 (2) (2011) 244–253.
[42] A. Orozco-Duque, J. Bustamante, G. Castellanos-Dominguez, Semi-supervised clustering of fractionated electrograms for electroanatomical atrial mapping, Biomed. Eng. Online 15 (44) (2016).
[43] R.J. Hunter, I. Diab, M. Tayebjee, L. Richmond, S. Sporton, M.J. Earley, R.J. Schilling, Characterization of fractionated atrial electrograms critical for maintenance of atrial fibrillation: a randomized, controlled trial of ablation strategies (the CFAE AF trial), Circulation: Arrhythm. Electrophysiol. 4 (5) (2011) 622–629.
[44] J. Kalifa, K. Tanaka, A.V. Zaitsev, M. Warren, R. Vaidyanathan, D. Auerbach, S. Pandit, K.L. Vikstrom, R. Ploutz-Snyder, A. Talkachou, F. Atienza, G. Guiraudon, J. Jalife, O. Berenfeld, Mechanisms of wave fractionation at boundaries of high-frequency excitation in the posterior left atrium of the isolated sheep heart during atrial fibrillation, Circulation 113 (5) (2006) 626–633.