

RESEARCH

Open Access



# Semi-supervised clustering of fractionated electrograms for electroanatomical atrial mapping

Andres Orozco-Duque<sup>1,2\*</sup>, John Bustamante<sup>1</sup> and German Castellanos-Dominguez<sup>3</sup>

\*Correspondence:

andres.oroazco@upb.edu.co;  
andresoroazco@itm.edu.co

<sup>2</sup> GI2B, Instituto Tecnológico  
Metropolitano, Medellín,  
Colombia

Full list of author information  
is available at the end of the  
article

## Abstract

**Background:** Electrogram-guided ablation procedures have been proposed as an alternative strategy consisting of either mapping and ablating focal sources or targeting complex fractionated electrograms in atrial fibrillation (AF). However, the incomplete understanding of the mechanism of AF makes difficult the decision of detecting the target sites. To date, feature extraction from electrograms is carried out mostly based on the time-domain morphology analysis and non-linear features. However, their combination has been reported to achieve better performance. Besides, most of the inferring approaches applied for identifying the levels of fractionation are supervised, which lack of an objective description of fractionation. This aspect complicates their application on EGM-guided ablation procedures.

**Methods:** This work proposes a semi-supervised clustering method of four levels of fractionation. In particular, we make use of the spectral clustering that groups a set of widely used features extracted from atrial electrograms. We also introduce a new atrial-deflection-based feature to quantify the fractionated activity. Further, based on the sequential forward selection, we find the optimal subset that provides the highest performance in terms of the cluster validation. The method is tested on external validation of a labeled database. The generalization ability of the proposed training approach is tested to aid semi-supervised learning on unlabeled dataset associated with anatomical information recorded from three patients.

**Results:** A joint set of four extracted features, based on two time-domain morphology analysis and two non-linear dynamics, are selected. To discriminate between four considered levels of fractionation, validation on a labeled database performs a suitable accuracy (77.6 %). Results show a congruence value of internal validation index among tested patients that is enough to reconstruct the patterns over the atria to located critical sites with the benefit of avoiding previous manual classification of AF types.

**Conclusions:** To the best knowledge of the authors, this is the first work reporting semi-supervised clustering for distinguishing patterns in fractionated electrograms. The proposed methodology provides high performance for the detection of unknown patterns associated with critical EGM morphologies. Particularly, obtained results of semi-supervised training show the advantage of demanding fewer labeled data and less training time without significantly compromising accuracy. This paper introduces a new method, providing an objective scheme that enables electro-physiologist to recognize the diverse EGM morphologies reliably.

**Keywords:** Atrial fibrillation, Electrogram-guided ablation, Feature extraction, Spectral clustering

## Background

Atrial Fibrillation (AF) implies that the electrical activity of the atria is highly disorganized, and any coherent mechanical contraction is missed. AF, which is the most common supraventricular arrhythmia, is associated with many cardiac conditions, including an increased risk of thromboembolic events, stroke and heart failure.

Catheter ablation has become an alternative to cure AF, and may avoid side effects of long term pharmacotherapy. Radiofrequency ablation treatment is the generation of tissue injuries which block propagation of electrical impulses to prevent the formation and maintenance of fibrillatory conduction. Catheters for radiofrequency ablation are guided inside the heart chambers via cardiac mapping systems [1].

Although electrical disconnection of the pulmonary veins remains the mainstream procedure of catheter ablation, patients with persistent AF demand more extensive ablation [2]. Recent approaches aim at guiding the ablation using electrical signals recorded inside the atria, called electrograms (EGM). These recordings are incorporated into an electroanatomical mapping system to visualize the 3D distribution of the electrical information through the anatomical atrial structure (electroanatomical atrial mapping – EAM). The main goal of EAM is to locate AF sources outside the region of pulmonary veins in cases of persistent AF.

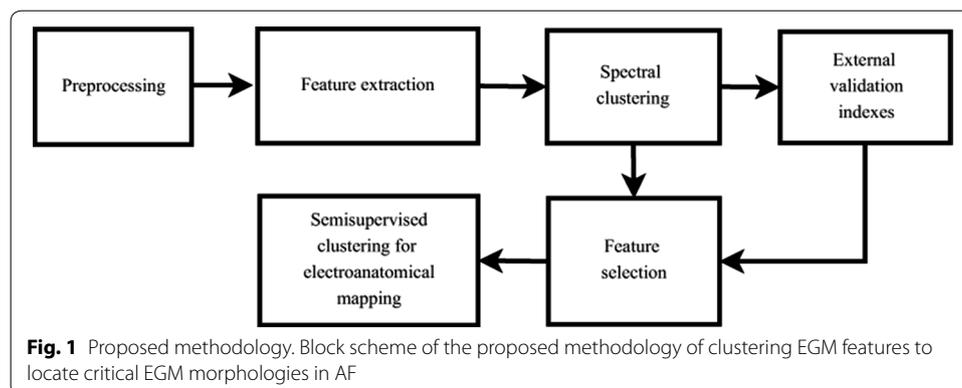
Even though the mechanism of AF remains unclear, some studies have shown that the EGM morphology during AF may be correlated with different conduction patterns, e.g., conduction blocks, slow conduction, a collision of activation waves or reentries [3]. In fact, areas rendering EGM recordings with remarked high-frequency content or chaotic patterns should be associated with AF [4, 5]. Thus, electrogram-guided ablation procedures have emerged as alternative strategy consisting of either mapping and ablating localized reentrant sources driving AF or targeting complex fractionated electrograms (CFAE) [6]. In accordance to [7], CFAE is formally defined as follow: (1) atrial electrograms that have fractionated electrograms composed of two deflections or more, 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. This inexact and wide-sense statement of CFAE makes the decision of selecting the target sites for ablation to be dependable on the expertise of the electrophysiologist, jeopardizing the effectivity of the CFAE ablation [8, 9]. To overcome these limitations, designation of different levels of fractionation (usually, between three and five) have been proposed based on the perturbation of baseline and the presence of continuous deflection [10, 11]. Every one of the fractionation levels and EGM morphologies remains not well described or is differently defined in the literature, making difficult their discrimination even for the electro-physicians. Therefore, there is a need for an objective scheme capable of distinguishing the diverse morphologies of EGM signals.

The extensive number of the feature extraction methods for the CFAE detection falls into the following categories: (i) features based on time-domain morphology analysis, e.g., measures of the cycle length [12], quantification of deflections [11], characterization

of baseline and wave similarity measure [13], among others; (ii) based on frequency analysis, e.g., dominant frequency and regularity index [14]; and (iii) based on nonlinear dynamics, such as Shannon entropy [15] and approximate entropy [16]. All these features aim at distinguishing each level of fractionation by building a single map encoding waveform differences of CFAE upon the anatomical structure of the atria [16]. Although most studied features have a simple implementation, they demand tuning of parameters that in practice should be heuristically fixed. Besides, because of the substantial stochastic behaviour of CFAE, the extraction of a unique feature has been proved to be not enough to identify all distinct substrates perpetuating the arrhythmia [17]. To date, feature extraction from complex fractionated electrograms is carried out based on mostly the time-domain morphology analysis and non-linear features instead of handling the entire waveform directly. However, we employ their combination that has been reported to achieve better performance [18].

On the other hand, most of the inferring approaches applied for identifying CFAE levels of fractionation are supervised. Examples are given in [19, 20], where sets of labeled signals must be used during the training process. Nonetheless, supervised learning is limited by the availability of marked CFAE, which in turn faces two restrictions: the lack of a standard for their objective description [17, 21, 22] and the fact that some of the CFAE properties may vary under the influence of different catheters or acquisition settings [23].

In order to overcome the above-described limitations, this work proposes a semi-supervised clustering method of four levels of fractionation. In particular, we use a spectral clustering that groups a set of widely used atrial EGM features extracted from complex fractionated electrograms. We also introduce a new atrial-deflection based feature quantifying the fractionated activity. Further, we select, from the input feature set, the optimal subset that yields the best performance. For purposes of evaluation of the proposed clustering method, we carry out training for two scenarios: (a) *External validation* using a labeled database with four different classes of atrial EGM. (b) *Internal validation* in a semi supervised fashion that employs the feature set extracted in the external validation, aiming to perform semi-supervised clustering on an unlabeled dataset recorded from three patients. The obtained results indicate that the proposed method is suitable for automatic identification of critical patterns in AF.



This work is organized as follows: in "Methods" section methods of feature extraction, spectral clustering, and feature selection are described. "Results of clustering" section carry out the results of experiments using both cases of validation on labeled and unlabeled databases. Lastly, we discuss all obtained results and provide conclusions in "Discussion" and "Conclusions" section, respectively.

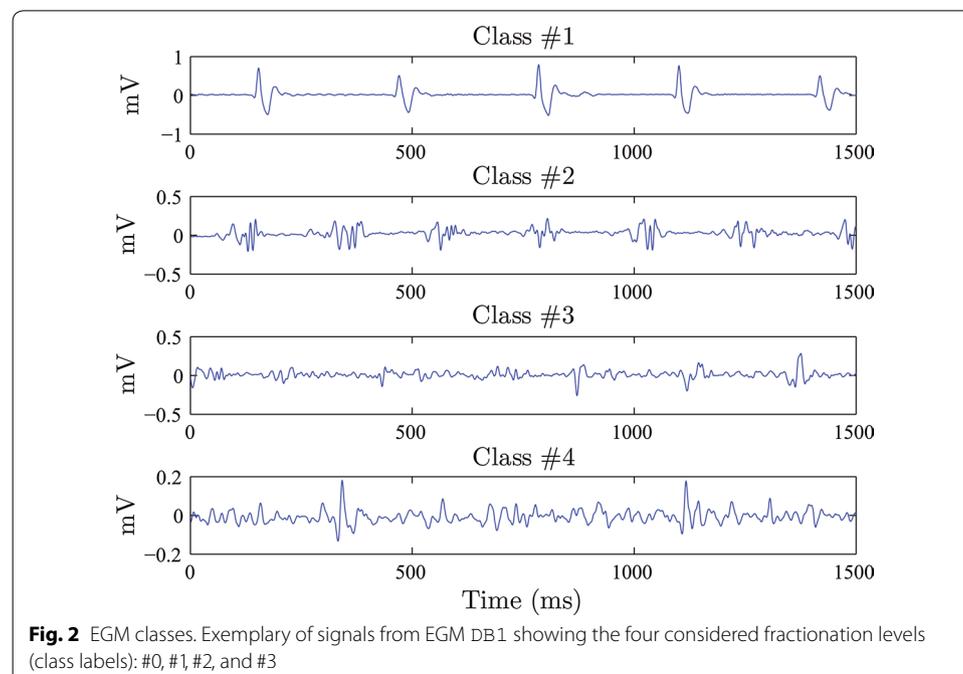
## Methods

With the aim at clustering EGM features for identification of ablation target areas, the proposed methodology comprises the following stages (see Fig. 1): (i) preprocessing, (ii) feature extraction, (iii) spectral clustering, (iv) feature selection, and (v) semi-supervised clustering for electro-anatomical mapping that displays the cluster labels in a color-coded overlaid on the reconstructed 3D atrial geometry of a patient.

### Tested EGM databases

#### Labeled EGM database (DB1)

This data collection holds 429 EGM recordings acquired from 11 AF patients, as established and reported in [20]. Intracardiac EGM recordings from a multipolar circular catheter were performed after pulmonary vein isolation with a sampling rate of 1.2 kHz. The database was independently annotated by two electrophysiologists, working at different centers, and with proved experience, according to predefined fractionation classes. Atrial EGM signals were checked visually and were labeled according the following fractionation levels (see Fig. 2): Non-fractionated EGM or level 0 (labeled as #0), mild, intermediate, and high (#1, #2, and #3, respectively). Besides, after visual inspection of the experts, the signals having the following particularities had been also sorted out: (i) signals with low quality with very low voltage, (ii) signals that are superimposed



on the ventricular far-field components, (iii) signals remain non-stationary over the whole five-seconds recording.

#### **Unlabeled EGM database (DB2)**

This collection was obtained at the Hamilton General Hospital.<sup>1</sup> Data were recorded from three patients having definite evidence of AF. The amount of 512 observations was acquired by sequential mapping during spontaneous AF before the circumferential ablation. Namely, 223, 88, is the average time between and 201 signals were recorded from the patients labeled as 1, 2, and 3 respectively. After ablation, all patients restored the sinus rhythm. For EGM acquisition, the circular mapping catheter scheme with 20 poles (2-6-4 mm spacing) was used by means of the EAM system Ensite™NavX™ (St. Jude Medical™). The catheter remained stationary during four seconds at each observation point. The data were acquired with a sampling rate of 2034.5 Hz. Besides the electrical data, the information about the anatomical model of the left atrial, acquired by the NavX™, were captured. The vertices and polygons to build the mesh that represent the atrial anatomic were also available. Additionally, the system provided the position of the electrode where every EGM was acquired. These information are used to construct an electro-anatomical map of the atrium for each patient.

#### **Feature extraction from electrogram morphology analysis**

To investigate the anatomic distribution of critical sources in patients with AF, several objective time-based measures are frequently performed, which essentially evaluate the salient organizational properties of the single atrial EGM recordings. Here, the following measures are considered (see Fig. 3):

- *Electrogram deflection time.* Deflections are those perturbations of the EGM baseline having the peak to peak amplitude greater than a given sensibility threshold,  $\epsilon_s \in \mathbb{R}^+$ . At the same time, the interval between adjacent peaks should last less than a pre-defined deflection width,  $\epsilon_w \in \mathbb{R}^+$ . Algorithm 1 computes a single vector of time deflections,  $\zeta \in \mathbb{R}^{n_d}$ , based on maxima and minima detection computed from the EGM signal.
- *Fractionation interval.* This parameter measures the period between two consecutive deflections (detected within the time range  $\zeta(j+1) - \zeta(j)$ ) which must be larger than the defined refractory period  $\epsilon_r \in \mathbb{R}^+$ .
- *Complex fractionated interval.* This interval covers uninterrupted electrical activity having consecutive deflection time values shorter than the effective refractory period of the atrial myocardium (70 ms [11]). Besides, all included deflections must exceed 20 % of the amplitude of the highest peak to peak deflection measured over the whole atrial electrogram. Algorithm 2 computes the output vector  $\mathbf{z} \in \mathbb{R}^N$  that represent the segments with fractionated electrical activity (see Fig. 3a).
- *Segments of Local Activation Waves (LAW).* This  $p$ -samples window holds all events of the local depolarization and is centered on the local atrial activation times (see Fig. 3b, c). For the LAW calculation, each measured atrial electrogram is filtered by

<sup>1</sup> <http://www.phri.ca/>.

a digital, zero-phase, third-order Butterworth filter with passband between 40 and 250 Hz as proposed in [24]. Algorithm 3 performs detection of LAW windows.

Consequently, the following features are extracted from the time-based measurements:

- *Complex fractionated electrogram (CFE) index*,  $\xi_1 \in \mathbb{R}^+$ , is the average time between fractionation intervals.
- *Fractionated activity*,  $\xi_2 \in \mathbb{R}^+$  describes the proportion of each EGM signal holding fractionated electrical activity, and is calculated by fixing the time instants when the sign of the envelope changes (i.e.,  $\mathbf{z} \neq 0$ ). Algorithm 2 computes the envelope  $\mathbf{z}$  of the input signal  $\mathbf{x}$ .
- *Variability of segments with fractionated electrical activity*,  $\xi_3 \in \mathbb{R}^+$  is the standard deviation of the width measured for the segments with fractionated electrical activity,  $\mathbf{w}$ , (see Algorithm 2).
- *Deflection-LAW ratio*,  $\xi_4 \in \mathbb{R}^+$ , is defined by the ratio  $\xi_4 = n_d/n_w$ , where  $n_d$  and  $n_w$  are computed from Algorithms 1 and 3, respectively.
- *Similitude index*,  $\xi_5 \in \mathbb{R}^+$ , is a wave-morphological resemblance between different local activation waves, quantifying the EGM regularity based on the degree of the LAW repeatability [13]. This index is defined as follows:

$$\xi_5 = \frac{2}{(n_w - 1)} \mathbb{E} \left\{ \sum_{j=1}^{n_w} \Theta(\epsilon - \arccos(\mathbf{s}_i, \mathbf{s}_j)) : \forall i = 1, \dots, n_w \right\} \quad (1)$$

where  $\Theta$  is the Heaviside function [25],  $\epsilon$  is a threshold adjusted to 0.8, and  $\mathbf{s}_i$  is the  $i$ -th detected LAW.

- *Dominant frequency index*,  $\xi_6 \in \mathbb{R}^+$ . This spectral component is inversely proportional to the cycle length. The dominant frequency is computed from the envelope  $g$  (see Algorithm 3) as the maximum peak of the Fast Fourier Transform power spectrum smoothed by the Hamming window.

---

**Algorithm 1** Detection of deflection times.

---

**Input:**  $\mathbf{x} = [x(1), \dots, x(N)]$  – the observations at sample time  $n$ ;

**Input:**  $\epsilon_w$  : deflection width,  $\epsilon_s$  : sensibility threshold

compute  $\mathbf{y} = -d\mathbf{x}/dt$ ,

**for**  $n = 1, \dots, N$  **do**

**if** Positive-to-negative Zero Crossing in  $y(n) \in \mathbf{y}$  **then**

$j = j + 1$ ; time counter

$p_1(j) = n$ : Fix the time instant at which every maximum value is detected;

**if** Negative-to-positive Zero Crossing in  $y(n)$  **then**

$p_2(j) = n$ : Fix the time instant at which every minimum value is detected

**for all**  $j$  in  $p_1(j)$  **do**

    Compute  $f(j)$  holding the time interval between  $p_1(j)$  and  $p_2(j)$

**if**  $f(j) < \epsilon_w$  **then**

        Reject  $p_1(j)$  and  $p_2(j)$ ;

$v(j) = x(p_1(j)) - x(p_2(j))$ : Compute the peak to peak voltage for  $j$ -th deflection

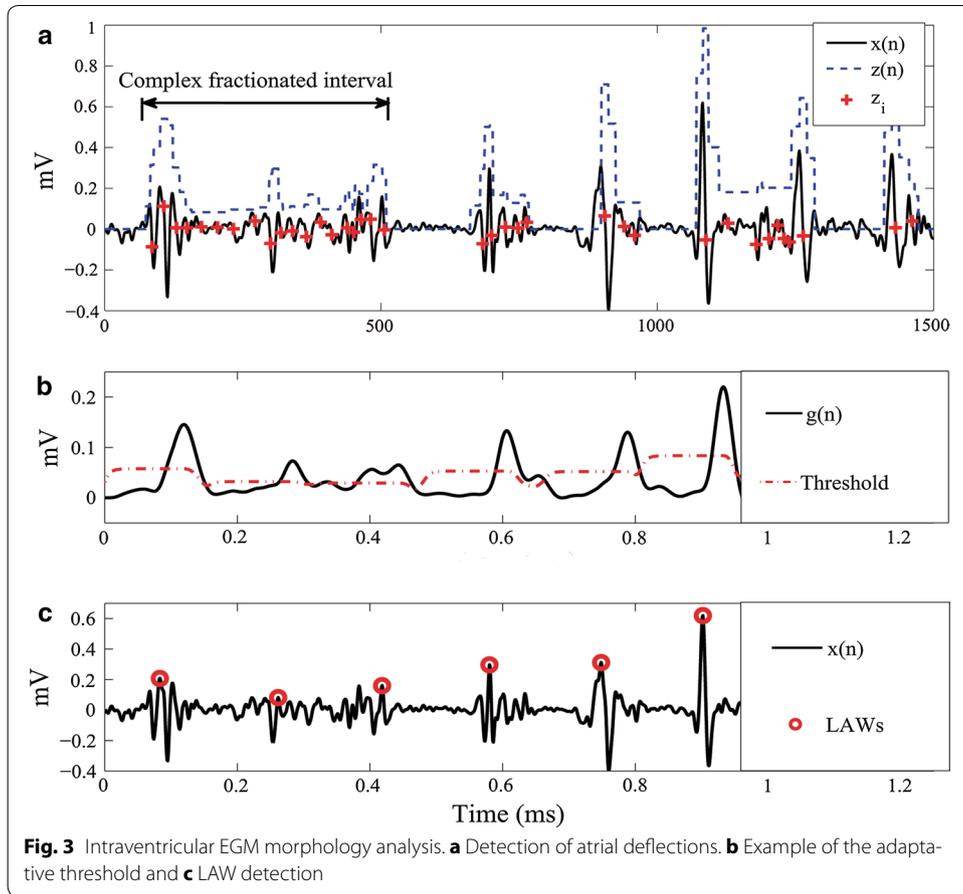
**if**  $v(j) < \epsilon_s$  **then**

        Reject  $p_1(j)$ ,  $p_2(j)$ , and  $v(j)$ ;

    Compute  $\zeta(j) = |p_1(j) + (p_2(j) - (p_1(j)))|/2$ : Fix each deflection time

**Output:** vectors  $\zeta, \mathbf{v}$ .

---



### Non-linear feature extraction from electrograms

Here, based on the non-linear dynamic theory, we also extract the following two non-linear features:

- The approximate entropy,  $\xi_7 \in \mathbb{R}^+$ , defined by the difference equation:

$$\xi_7 = \Phi^m(r) - \Phi^{m-1}(r) \tag{2}$$

where  $m \in \mathbb{N}$  is the embedded dimension,  $r \in \mathbb{R}^+$  is a threshold of minimum tolerance, ranging from 0.1 to 0.5 times the standard deviation of the signal. Here, the real-value functional  $\Phi^m(r) \in \mathbb{R}^+$  is computed as:

$$\Phi^m(r) = \mathbb{E} \left\{ \log \left( \mathbb{E} \left\{ \Theta(r - d(\mathbf{x}_i^m, \mathbf{x}_j^m)r) : \forall j = 1, \dots, N - m + 1 \right\} \right) : \forall i \neq j \right\}$$

where notation  $\mathbb{E}\{\cdot\}$  stands for the expectation operator;  $\Theta \in [0, 1]$  is the Heaviside function applied to the used measure of similarity between each couple of EGM lagged versions,  $\mathbf{x}_i^m$  and  $\mathbf{x}_j^m$ :

$$d(\mathbf{x}_i^m, \mathbf{x}_j^m) = \max_{k=1,2,\dots,m} (|x(i+k-1) - x(j+k-1)|),$$

where either lagged vector  $\mathbf{x}_k^m = [x(k), \dots, x(k-m+1)]$  (with  $\mathbf{x}_k^m \in \mathbb{R}^m$ ) holds the  $m$  consecutive samples of the original signal,  $\mathbf{x}$ , starting at the  $i$ -th time instant.

- The multifractal  $h$ -fluctuation index [26],  $\xi_8 \in \mathbb{R}$ , is defined as the power of the second order backward difference of the generalized Hurst exponent  $h(q) \in \mathbb{R}$  as follows [26]:

$$\xi_8 = \frac{1}{2^{|q_{\max}| - 2}} \sum_{q=q_{\min}+2}^{q_{\max}} (h(q) - 2h(q-1) + h(q-2))^2 \tag{3}$$

where  $q \in \mathbb{N}$  is the order for evaluating the partition function, providing  $q_{\min} < 0, q_{\max} > 0$  and  $|q_{\min}| = |q_{\max}|$ ;  $q_{\min}$  is the minimum negative order  $q$ , and  $q_{\max}$  is the maximum positive order  $q$  used in the estimation of multi-fractal spectrum through the multi-fractal detrended fluctuation analysis.

---

**Algorithm 2** Detection of segments with fractionated electrical activity.

---

**Input:**  $\mathbf{x}$ ,  $\mathbf{v}$ , and  $\mathbf{v}$  calculated in Algorithm 1,

Make  $z(n) = 0, \forall n = 1, \dots, N$

**for all**  $j = 1, \dots, n_d$  **do**  
     **if**  $v(j) > (0.20 \cdot \max_{v_j}\{v(j)\})$  **then**  
         **for**  $n = \zeta(j-1), \dots, \zeta(j)$  **do**  
              $z(n) = v(j)$

    Compute  $\mathbf{w} \in \mathbb{R}^{n_f}$  holding the  $n_f$  segment duration values of  $z(n)$  where  $z(n) \neq 0$ .

**Output:** vectors  $\mathbf{z}$  and  $\mathbf{w}$

---

Consequently, we extract  $D = 8$  features for identification and localization of critical sources in AF, resulting in the atrial EGM feature point  $\xi = [\xi_1, \dots, \xi_D]$  that describes each electrogram.

**EGM feature clustering for identification of ablation target areas**

**Spectral clustering of atrial EGM features**

Let  $\mathbf{E} \in \mathbb{R}^{M \times D}$  be an input data matrix holding  $M$  objects and  $D$  features, where each row  $\{\xi_i \in \mathbb{R}^D : i = 1, \dots, M\}$  denotes one single data point. The goal of clustering is to divide the data into different groups, where samples gathered within the same group are similar to each other. To discover the main topological relationships among data points, spectral clustering-based approaches build from  $\mathbf{E}$  a weighted graph representation  $\mathcal{G}(\mathbf{E}, \mathbf{K})$ , where each object point,  $\xi \subseteq \mathbf{E}$ , is a vertex or node and  $\mathbf{K} \in \mathbb{R}^{M \times M}$  is a similarity (affinity) matrix encoding all associations between graph nodes. In turn, each element of the similarity matrix,  $k_{ij} \subseteq \mathbf{K}$ , corresponding to the edge weight between  $\xi_i$  and  $\xi_j$ , is commonly defined as follows [27]:  $k_{ij} = \mathcal{K}(\xi_i, \xi_j; \sigma)$ ,  $k_{ij} \in \mathbb{R}^+$ , where function

$$\mathcal{K}(\xi_i, \xi_j; \sigma) = \exp\left(-\|\xi_i - \xi_j\|_2^2 / 2\sigma^2\right)$$

is the Gaussian kernel, and  $\sigma \in \mathbb{R}^+$  is the kernel bandwidth. Notation  $\|\cdot\|_2$  stands for the  $L_2$ -norm. Although there are many available kernels (like the Laplacian or polynomial ones), the Gaussian function has the advantages of finding Hilbert spaces with universal approximating capability and of being mathematically tractable.

---

**Algorithm 3** Detection of LAW windows

---

**Input:** vector  $\mathbf{x}$   
 Compute envelope  $\mathbf{g}(n)$  by submitting the input signal  $x$  to a 40 – 250 Hz band-pass filter to obtain  $x_f$ . Then, smooth  $|\mathbf{x}_f|$  using a Butterworth low-pass filter with 20 Hz-cut-off;  
**for all**  $n \in N$  **do**  
     Calculate the adaptive threshold  $\epsilon_a$  based on Pan and Tompkins algorithm;  
     Compute  $n_w$ , the number of time intervals where  $\mathbf{g}$  is greater than  $\epsilon_a$ .  
     Compute  $\mathbf{1} \in \mathbb{R}^{n_w}$  holding the time instants where  $\mathbf{g}$  is greater than  $\epsilon_a$ .  
     Compute  $\mathbf{2} \in \mathbb{R}^{n_w}$  holding the time instants where  $\mathbf{g}$  is less than  $\epsilon_a$ .  
**for**  $i = 1, \dots, n_w$  **do**  
     Fix the time instants of LAW centers:  $\sigma(i) \leftarrow$  the time instant where  $\mathbf{x}$  is the max in the interval  $[\gamma_1(i), \gamma_2(i)]$ .  
**Output:**  $\sigma(i)$

---

Hence, the clustering task now relies on the conventional graph cut problem that aims at partitioning a set of vertices  $\mathcal{V} \in \mathcal{E}$  into  $C \in \mathbb{N}$  disjoint subsets  $\mathcal{V}_c$ , so that  $\mathcal{V} = \cup_{c=1}^C \mathcal{V}_c$  and  $\mathcal{V}_{c'} \cap \mathcal{V}_c = \emptyset, \forall c' \neq c$ . Since the graph-cut approaches demand high computational power, relaxation of the clustering optimization problem has been developed based on the spectral graph analysis [28]. So, spectral clustering-based methods decompose the input data  $\mathcal{E}$  into  $C$  disjoint subsets by using both spectral information and orthogonal transformations of  $\mathbf{K}$ . Algorithm 4 describes the well-known solution of the cut problem (termed *NCut*).

---

**Algorithm 4** Basic Spectral Clustering

---

**Input:**  $\mathbf{E} \in \mathbb{R}^{M \times D}, \mathbf{K} \in \mathbb{R}^{M \times M}$ ,  
**Input:**  $C \in \mathbb{M}$  number of clusters determined by the the structure of the eigenvectors.  
 Compute the diagonal degree matrix  $\mathbf{W} \in \mathbb{R}^{M \times M}$  holding elements  $w_{ii} = \sum_{j \in M} k_{ij}$ .  
 Calculate the normalized Laplacian matrix:  $\mathbf{L} = \mathbf{W}^{-1/2} \mathbf{K} \mathbf{W}^{-1/2}$ .  
 From  $\mathbf{L}$ , calculate the sets of eigenvalues,  $\{\lambda_i \in \mathbb{R}^+\}$ , and eigenvectors  $\{\mathbf{v}_i \in \mathbb{R}^M : i=1, \dots, M\}$ , such that  $\{\lambda_i \geq \lambda_j\}, \forall 1 < i < j < M$ .  
 Construct  $\mathbf{V} = [\mathbf{v}_1; \dots; \mathbf{v}_C]$  by stack the  $C \leq M$  first eigenvectors into the columns of  $\mathbf{V} \in \mathbb{R}^{M \times C}$ .  
 Provided each row of  $\mathbf{V}$  with dimension  $\mathbb{R}^C$ , cluster all matrix rows into  $C$  groups.  
 Assign the original point  $\mathbf{x}_i$  to cluster  $c$ , if and only if, the  $i$ -th row of the matrix  $\mathbf{V}$  is allocated to the cluster  $c$ .  
**Output:**  $\mathcal{V} = \{\mathcal{V}_c : c=1, \dots, C\}$ .

---

**Selection of the optimal EGM feature set**

Given an input feature matrix  $\mathbf{E} \in \mathbb{R}^{M=D}$ , the aim of the feature selection stage is to find the optimal subset  $\mathcal{E}^*$  that holds  $D' < D$  selected features and provides the highest performance, measured in terms of the cluster validation. For searching  $\mathcal{E}^*$ , we implemented the Sequential Forward Selection (SFS). At the first iteration, the SFS selects the feature with best performance. In the next iteration, all candidate subsets combining two features (including the one selected before) are evaluated, and so on. This procedure is carried out iteratively by adding all previously selected features and ceases when the following stopping criterion supplies the minimum value:

$$\mu_{sc} = -(\mu_1 - \mu_2), \tag{4}$$

where  $\mu_{sc} \in \mathbb{R}[-1, 1]$ , is the trade-off between the following two indexes of clustering performance:  $\mu_1 \in \mathbb{R}[0, 1]$  is the Adjusted Rand Index that is an external counter checking whether the inferred labels and a set of external labels resemble the same structure [29], and  $\mu_2 \in \mathbb{R}[0, 1]$  is the equivalence mismatch distance that counts all pairs of labels,

which have different assignation. Additional explanation about both cluster validation indexes is given in [Appendix](#).

### Results of clustering

For purposes of evaluation of the clustering quality, we carry out training using the selected feature set in two cases: *a) External validation* using a labeled database with four different classes of atrial EGM. *b) Semi-supervised clustering* that employs a small amount of labeled data, used in the first training case, to aid semi-supervised clustering on unlabeled dataset, associated with anatomical data, performed separately for each patient.

#### Parameter setting for feature estimation

In the beginning, each acquired EGM,  $x \in \mathbb{R}^N$ , is firstly submitted to a 30–500 Hz band-pass filter and then passed through a 60 Hz notch filter, being  $N = 6000$  the signal length. Both procedures are performed by means of the NavX™ system.

In order to accomplish the feature extraction stage from the EGM morphology analysis, we detect deflections fixing  $\epsilon_w = 20$  ms as recommended in [11]. The parameter  $\epsilon_s$  is set differently for each database: For DB1,  $\epsilon_s = 0.01$  of the normalized recording amplitude. For DB2, we fix  $\epsilon_s = 0.05$  mV since there is just one patient under examination, making unnecessary the normalization of the recordings. Based on the detected set of deflections, the CFE index  $\xi_1$  is calculated assuming  $\epsilon_r = 30$  ms. Besides, the computation of similitude index  $\xi_5$  is carried out adjusting  $p = 90$  ms [13].

For the extraction of the non-linear feature,  $\xi_7$ , the following parameters are fixed, as suggested in [16]: Embedded dimension  $m = 3$  and a threshold  $r$  equal to 0.38 times the standard deviation of the signal. As explained in [16], The optimal value of  $r$  and  $m$  is the trade-off between the interclass percentile distance that minimizes the scatter in each class and the interclass minimum-maximum distance that maximizes the distances between the feature measures of the classes. Lastly, calculation of  $\xi_8$  is performed from the multifractal detrend fluctuation analysis, where the values  $q_{\min} = -5$  and  $q_{\max} = 5$  are fixed heuristically.

#### Clustering-based feature selection

We carry out supervised spectral clustering on DB1 to discriminate between the four levels of fractionation ( $C = 4$ ). As given in [30], we set the kernel parameter  $\sigma$  using the

**Table 1 The effect of the choice of features on spectral clustering**

Optimal feature set								$\mu_1$	$\mu_2$	$\mu_{sc}$
$\xi_2$								0.459	0.225	-0.234
$\xi_2$	$\xi_8$							0.514	0.197	-0.317
$\xi_2$	$\xi_8$	$\xi_7$						0.491	0.205	-0.286
$\xi_2$	$\xi_8$	$\xi_7$	$\xi_5$					0.521	0.193	-0.327*
$\xi_2$	$\xi_8$	$\xi_7$	$\xi_5$	$\xi_1$				0.495	0.206	-0.286
$\xi_2$	$\xi_8$	$\xi_7$	$\xi_5$	$\xi_1$	$\xi_4$			0.492	0.235	-0.257
$\xi_2$	$\xi_8$	$\xi_7$	$\xi_5$	$\xi_1$	$\xi_4$	$\xi_3$		0.483	0.235	-0.248
$\xi_2$	$\xi_8$	$\xi_7$	$\xi_5$	$\xi_1$	$\xi_4$	$\xi_3$	$\xi_6$	0.450	0.243	-0.207

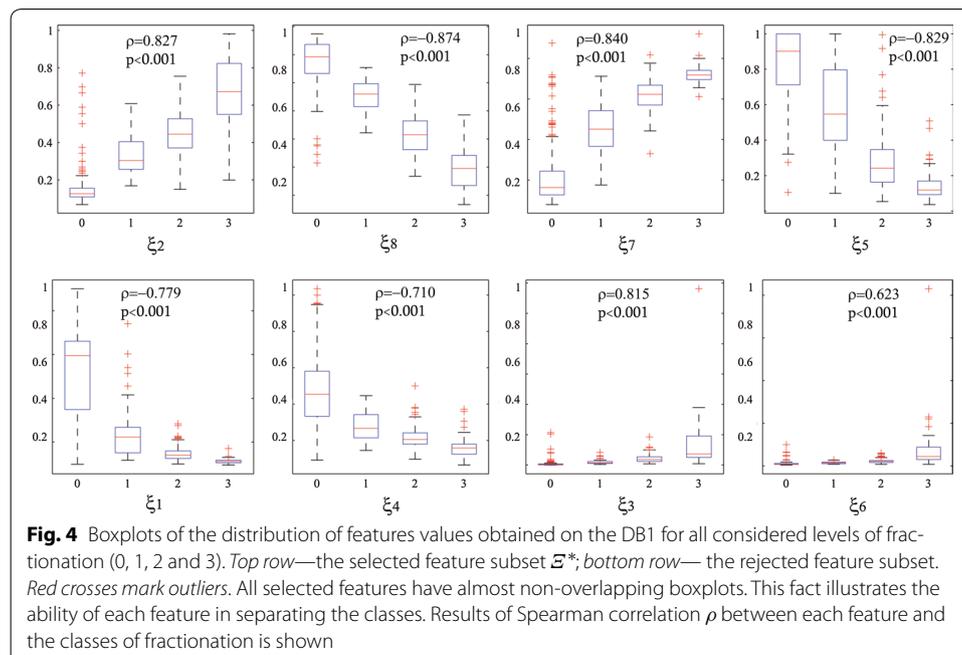
Notation (\*) points out on the selected feature subset,  $\mathcal{E}^*$ , that reaches the lowest value of  $\mu_{sc}$

tuning method based on the maximization of the transformed data variance as function of the scaling parameter. Further, we complete the feature selection stage that uses all available labels. As shown in Table 1, the most relevant feature is  $\xi_2$ , while the selected optimal feature subset is  $\mathcal{E}^* = \{\xi_2, \xi_8, \xi_7, \xi_5\}$  which is the one that reaches the best trade-off value of the minimizing cost function  $\mu_{sc}$ .

Figure 4 displays the boxplot diagrams that include the median values and the inter-quartile ranges of each feature, calculated for all considered levels of fractionation. In the top row, the boxplot diagrams of the selected feature subset  $\mathcal{E}^*$  illustrate the ability of each feature in separating the classes of fractionation levels. All selected features have almost non-overlapping boxplots. This fact favors the distinction of the fractionation levels, since their medians are separated enough from each other. In fact, the results of the carried out Spearman correlation test confirm this assumption. However, a detailed visual inspection of the diagrams shows that the class labeled as #0 (that is, non-fractionated EGM) has the highest number of outliers. By contrast, the class #1 (mild fractionation) holds no outliers at all. In the bottom row, the displayed boxplot diagrams are clearly overlapped, causing that this feature subset to be rejected. Note the poor performance achieved by the features  $\xi_3$  (Variability of complex fractionated segments) and  $\xi_6$  (dominant frequency index).

#### Clustering performance for the external validation

Here, experiments were focused on comparing the clustering results produced by the criterion of feature selection, proposed in Eq. (4), with the ground truth labels provided by DB1. Thus, Spectral clustering was carried out on the selected subset of relevant features,  $\mathcal{E}^*$ . For the sake of comparison, we did the same for the complete EGM feature set  $\mathcal{E}$ , for the selected morphology-base features, for the selected non-linear features

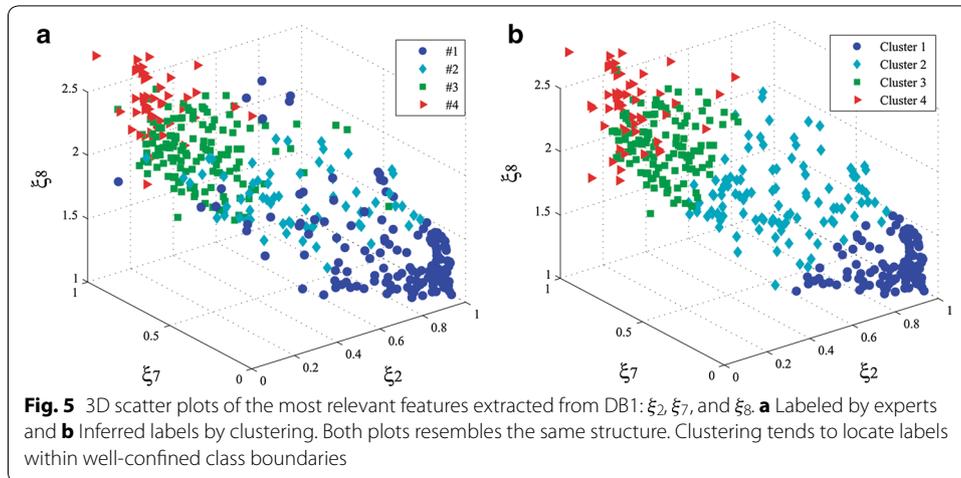


**Table 2 Performed external validating measures of spectral clustering on the labeled ground truth data DB1**

<b>(a) Performance using <math>\mathcal{E}</math></b>				
<b>Acc.</b>	<b>Spec.</b>			<b>Sens.</b>
47.55	93.47			84.31
	53.34			76.00
	85.05			11.48
	100.0			1.88
<b>(b) Performance using <math>\mathcal{E}^*</math></b>				
<b>Acc.</b>	<b>Spec.</b>			<b>Sens.</b>
77.62	98.91			71.24
	85.87			78.66
	88.61			84.45
	97.07			75.47
<b>(c) Confusion matrix using <math>\mathcal{E}^*</math></b>				
	<b>#0</b>	<b>#1</b>	<b>#2</b>	<b>#3</b>
<b>#0</b>	113	36	1	3
<b>#1</b>	1	66	8	0
<b>#2</b>	0	18	115	15
<b>#3</b>	0	0	14	39
<b>(d) Accuracy of different sets</b>				
<b>Morphology-based</b>	<b>Non-linear</b>		<b>Raw waveform</b>	
69.46 %	70.86 %		36.6 %	

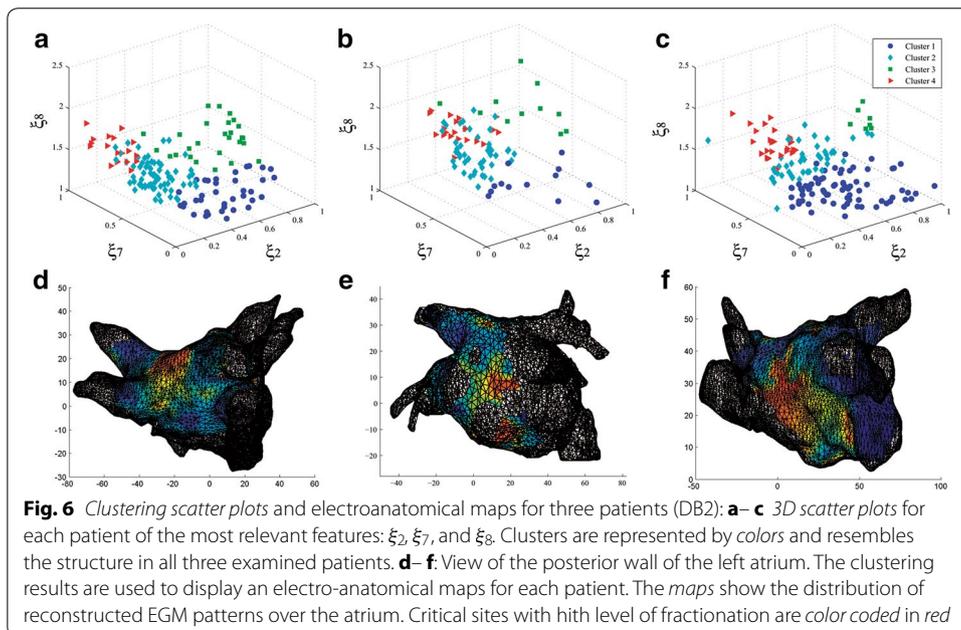
and for the raw-waveform. Table 2 shows the achieved clustering performance measured in terms of sensitivity, specificity, and accuracy for each level of fractionation of DB1. All these performance measures were calculated by direct comparison between the labels provided by an expert and the labels yielded by the spectral clustering technique. Table 2a and b show the computed measures for spectral clustering on subsets  $\mathcal{E}$  and  $\mathcal{E}^*$ , respectively. As it can be seen, the use of the latter features improves the detection performance remarkably. It is worth noting that the former set  $\mathcal{E}$  includes the CFE index,  $\xi_1$ , defection ratio,  $\xi_4$ , variability of complex fractionated segments,  $\xi_3$ , and dominant frequency index,  $\xi_6$ ; all these features are related to features extracted from EGM morphology analysis.

On the other hand, the selected feature set  $\mathcal{E}^*$  still supplies low sensitivity for the classes labeled as #0 and #3, as shown in the corresponding confusion matrix of Table 2(c). For getting a better insight into this issue, Fig. 5 displays 3D scatter plots allowing the visualization of the multivariate features  $\xi_2$ ,  $\xi_7$ , and  $\xi_8$ . As it can be seen in Fig. 5a, which shows the labels assigned by the expert panel, the expert’s markers tend to be more scattered just for the classes #0 and #3. Apparently, all these spread points are not taken into account by the clustering procedure, as this tends to locate labels within well-confined class boundaries, as shown in Fig. 5b.



### Semi-supervised clustering of unlabeled clinical data

We apply transductive learning to infer the correct labels for the unlabeled samples acquired from the same patient (see DB2), where the cluster assumption holds. Consequently, we assume that unlabeled data tend to form groups clearly separable so that the points of each partition should share one label. The detected EGM classes are handled for visualizing, in a color-coded map, the distribution of the EGM morphologies over the atria in the 3D mesh of the atrium. Thus, the electrophysiologists can locate more accurately the basic EGM classes that have highly fragmented morphologies. To this end, we use just the selected feature set,  $\mathcal{E}^*$ , that had been inferred by the above-supervised clustering procedure for the labeled data DB1. For the sake of visual inspection, the first row of Fig. 6 displays the estimated 3D scatter plots using the most relevant features



( $\xi_2$ ,  $\xi_7$ , and  $\xi_8$ ). As seen in Fig. 6a–c, the location of the clusters resembles the structure in all three examined patients.

To make clear the contribution of this transductive approach, we compare the inferred clusters by quantifying the similarity between partitions achieved for each case of training, supervised and semi-supervised. To this end, the Silhouette Index that ranges within the real-valued interval  $[-1, 1]$  can be calculated as the ratio of the intercluster cohesion versus to the intracluster separation [31]. Silhouette Index estimates the clustering consistency for each patient, fixing the number of fractionated levels as  $C = 4$ . The calculated Silhouette Index is 0.471 for patient 1, 0.481 for patient 2 and 0.469 for patient 3, while the same score is 0.57 for DB1, meaning that all carried out partitions tend to be similar in terms of cluster consistency.

The bottom row of Fig. 6 shows three EAM in which all EGM patterns are display over a mesh of the left atrium. The mesh is reconstructed using the anatomical information. EAM allows displaying on color scales the distribution of different EGM classes by their anatomical location at the atrial surface. In this work, the labels assigned by spectral clustering are used for setting the color scale regarding the level of fractionation. The color ranges from the blue that corresponds to non-fractionated signals to the red color standing for the highest level of fractionation. The obtained electroanatomical atrial mapping enables electro-physicians to recognize the location of diverse EGM morphologies on the atrial surface.

## Discussion

In this work, we propose a novel method to construct an semi-supervised-clustering-based electroanatomical map to display the distribution of EGM patterns in the atrial surface. The proposed methodology of training includes the use of a reduced set of features extracted from electrograms, providing a suitable performance. So, our method discriminates four EGM classes and benefits the ablation therapy since it provides an objective scheme that enables electro-physiologist to recognize the diverse EGM morphologies reliably. In accordance with the results obtained in the above section, the following findings are worth mentioning:

- In medical practice, the intracavitary mapping techniques are employed for the ablation in patients suffering from AF. Nevertheless, electrophysiologists must target the critical regions as accurately as possible, aiming to increase the effectiveness of radiofrequency ablation therapy. However, there is an incomplete understanding of the mechanism ruling the AF. Thus, the fractionation levels and EGM morphologies are often vaguely described or differently defined in the professional literature, making very hard their discrimination even for the electro-physicians. This aspect also complicates the automated training. As a result, there a very few available EGM datasets with proper labels. Just, our proposed approach is based on semisupervised clustering when unlabeled data are employed in conjunction with a small amount of labeled data.
- For localization of critical AF drivers in patients with AF, the baseline feature extraction method is grounded on the electrogram morphology analysis. Here, we consider the following five atrial-deflection based features: Complex fractionated electrogram

index, fractionated activity, variability, deflection-law ratio, similitude index, and the Dominant frequency index. Two non-linear features are also extracted: Approximate entropy and  $h$ -fluctuation index. We also carried out feature selection of the optimal subset, yielding the best possible performance of the clustering. Here, the sequential forward selection is implemented, for which we propose a stopping criterion based on the clustering performance. As a result, the following features are selected, ranked by relevance: fractionated activity  $\xi_2$ ,  $h$ -fluctuation index  $\xi_8$ , approximate entropy  $\xi_7$ , and similitude index  $\xi_5$ . The first feature, fractionated activity index,  $\xi_2$ , is a time-based measure relating to atrial deflections and describes the proportion of EGM signal holding all segments with fractionated electrical activity. Even though there are other similar indexes reported in literature [10, 32], they require some heuristical thresholds that in practice demand a considerable effort to tune. By contrast, the  $\xi_2$  is adjusted according to the effective refractory period of the atrial myocardium, which supplies more reliable physiological information. On the other hand, the following features extracted from electrogram morphology analysis were rejected: the complex fractionated electrogram index  $\xi_1$ , the deflection ratio  $\xi_4$ , the variability of complex fractionated segments  $\xi_3$ , and the dominant frequency index  $\xi_6$ . Furthermore, the relevance of the baseline CFE index  $\xi_1$  (termed as CFE-mean in the NavX™ system), which has been widely used in some commercial equipments, appears to be very poor, at least in terms of distinguishing among fractionation levels. Clinical studies report that it is unclear whether CFE-index is related with atrial substrates [17]. These results may be explained in the light of the highly non-stationary behavior of the EGM signals, making it difficult to achieve a confident estimation of the time-domain measures performing only the electrogram morphology analysis.

- Even that features extraction from fractionated electrograms is carried out based on mostly the time-domain morphology analysis [11, 33] and non-linear features [15, 16, 34] instead of handling the entire waveform directly, we employ their combination that has been reported to achieve better performance [10, 20]. Our performed training results on the tested database clearly support this statement [see Table 2(d)]: selected morphology-based feature set (69.46 %), selected non-linear set (70.86 %), and selected joint set (77.62 %). For the sake of comparison, we also tested the training using the waveform based input, reaching a very low performance (36.6 %). Obtained results show that the mixture of non-linear and morphology features can more efficiently encode the properties of AF patterns. These findings are consonant with clinical studies that had been carried out for simulation modeling [15] or animal [5] and human models [35], making the combination of EGM features a promising way to discriminate arrhythmogenic substrates.
- Atrial EGM signals are commonly labeled by three to five fractionation levels due to the influence of the baseline perturbation and continuous deflections [19]. For automating the labeling of ablation target areas, we make use of semi-supervised clustering into four levels of fractionation. Although there are several basic clustering methods, we employ the spectral clustering technique that provides two advantages: performing well with non-Gaussian clusters and totally automated the procedure of parameter settings. Another aspect of consideration is the generalization ability of the used semi-supervised clustering, because it does not make strong assumptions

on statistics of the classes. This latter property supplies adequate performance at small patient-specific EGM sets.

- To the best knowledge of the authors, the use of semi-supervised clustering for distinguishing among fractionated levels has not been discussed before. The primary goal of this approach is to make available an automatic training devoted to electro-anatomical atrial mapping, avoiding as much as possible the manual classification of AF types and reducing the dependence of prior knowledge about the statistics of the classes. Since manual AF labeling is subjective and time-consuming, it can be achievable for small databases. External validation using a labeled ground truth database with four different levels of fractionation achieved an accuracy of 77.6 %. This performance is comparable to the one (80.65 %) produced by the alternative supervised approach using a fuzzy decision tree in [20]. However, the supervised methods of classification, trained with short training datasets, tend to be biased due to the subjective labeling of AF types suffers from poorly described patterns and strong assumptions on statistics of the classes. This is an important property in this application due the lack of a standard definition of fractionated EGM. In fact, the generalization ability of the proposed training approach is tested to aid semi-supervised learning on unlabeled dataset recorded from three patients. The relevance of locating EGM patterns is encouraged by several studies pointing out that some particular fractionated morphologies are likely to represent drivers of AF [36]. Moreover, experimentation on isolated animal hearts has shown that the areas with highest fractionated EGM signals coexist in the periphery of the most rapid and less fractionated places [4, 37]. This fact may lead to the localization of AF sources and implies that the localization of different patterns, over the patient atrial surface, can become an adequate diagnostic support tool for locating target sites for ablation.
- The proposed methodology of training is devoted to automatic identification of different patterns in atrial EGM during AF. The commonly used systems to perform ablation (NavX system or Carto system) have a limited number of simultaneous EGM electrodes [11]. This fact implies that the EGM signals are asynchronous, and the reconstruction of action potential propagation around the whole atria is unfeasible. The proposed semi-supervised training allows inferring unknown patterns, which can be correlated with AF critical areas, so that it can improve the performance of the ablation therapy, even if the conventional mapping catheter is employed.
- Although electrical isolation of pulmonary veins is the mainstream ablation procedure for AF, CFAE ablation together with pulmonary vein isolation has attracted attention in reducing the long-term recurrence of AF [38]. Nevertheless, the latter ablation remains a debated issue due to the uncertainty of interpretation about many CFAE morphologies [36]. In this regard, the proposed semi-supervised mapping method can favor the use of EGM-guided ablation due to its ability for locating the distribution of different fractionated EGM patterns over the atrial for persistent AF patients. Therefore, the proposed method could be used in clinical studies to establish a relationship between EGM patterns and drivers that maintain AF, aiming to guide ablation procedures in patients with persistent AF.
- Lastly, we measure the computational complexity of the method in terms of processing time. The feature extraction step lasts 2 s for each signals. Provided a testing set

that holds 220 EGM signals (the average amount of signals for a mapping procedure), the spectral clustering lasts 0.56 s, and the mapping construction takes only 0.47 s. This time was calculated using MatLab 2013a in a PC with Windows 8 (64 bits), Core I7 processor and RAM of 6 GB. In total, the proposed training algorithm takes a short period so that the method can be employed for clinical purposes.

## Conclusions

This paper introduces a new method for semi-supervised clustering of fractionated electrograms, providing an objective tool for reliably locating the distribution of different fractionated EGM patterns over the atrial. The obtained electroanatomical atrial mapping enables electrophysiologist to locate the critical EGM patterns as accurately as possible, aiming to increase the effectiveness of radiofrequency ablation therapy for persistent AF patients.

Also, we introduce a new atrial-deflection based feature (termed fractionated activity) that does not demand any heuristical parameter tuning, providing an increased discrimination ability in comparison to the other state-of-the-art features. Furthermore, our carried out feature selection allows coming to the conclusion that some used in practice features (like the CFE index) have questionable effectiveness to localization of critical sources in patients with AF. Also, the use of semi-supervised clustering facilitates the automatic detection of fractionation classes with accuracy comparable to other similar results reported in the literature, avoiding the manual labeling of AF classes that is subjective and very time-consuming.

As the future work, the authors plan to improve the performance of the discussed semi-supervised clustering of features extracted from fractionated electrograms. Besides, a more detailed study should be carried out to discriminate different patterns over the atrial surface to be further associated with the fibrillatory conduction. We also plan to conduct clinical assessment of the effectiveness of the proposed method as a new electro-anatomical mapping tool to guide ablation procedures in AF.

## Abbreviations

AF: atrial fibrillation; EGM: electrograms; EAM: electro-anatomical atrial mapping; CFAE: complex fractionated atrial electrogram; LAW: local activation waves; CFE: complex fractionated electrogram; SFS: sequential forward selection; DB1: labeled EGM database; DB2: unlabeled EGM database.

## Authors' contributions

AOD and GCD participated in the design of the entire experiment research and helped to draft the manuscript. They also contributed to model implementation and interpreted the data results. JB structured medical background and helped in interpreting results of experiments. All authors read and approved the final manuscript.

## Author details

<sup>1</sup> Bioengineering Center, Universidad Pontificia Bolivariana, Medellin, Colombia. <sup>2</sup> GI2B, Instituto Tecnológico Metropolitano, Medellin, Colombia. <sup>3</sup> Signal Processing and Recognition Group, Universidad Nacional de Colombia - sede Manizales, Manizales, Colombia.

## Acknowledgements

We would like to acknowledge at the Institute of Biomedical Engineering at Karlsruhe Institute of Technology in Germany, and also Dr. Armin Luik from the Department of Cardiology and the Department of Internal Medicine at Karlsruhe City Hospital in Germany for providing databases of AF signals. We also appreciate the knowledge and data collected with the help of Dr. Carlos Morillo and his colleagues at the Population Health Research Institute, Hamilton Health Sciences, McMaster University, Hamilton, Ontario, Canada. This work has been supported by COLCIENCIAS—Republica de Colombia, project No. 121065741044, and by and ARTICA 1234 Project, Colombia.

## Competing interests

The authors declare that they have no competing interests.

### Appendix: Measures of cluster validation

The Adjusted Rand Index (ARI) is an external counter that checks whether the labels of the used clustering procedure,  $\mathcal{V}$ , and a set of external labels,  $\mathcal{V}'$ , resemble the same structure. ARI counts each pair-wise verification affiliating objects to the following subsets: Subset *a*) objects labeled in the same cluster of  $\mathcal{V}$  and  $\mathcal{V}'$ , *b*) objects labeled in the same cluster of  $\mathcal{V}$ , but in different clusters of  $\mathcal{V}'$ , *c*) objects labeled in the same cluster of  $\mathcal{V}'$ , but in different cluster of  $\mathcal{V}$ ; and *d*) objects labeled in the different cluster of both  $\mathcal{V}$  and  $\mathcal{V}'$ . Provided the above subsets, ARI is rated as follows [29]:

$$\mu_1 = \frac{\binom{M}{2}(a+d) - ((a+b)(a+c) + (c+d)(b+d))}{\binom{M}{2}^2 - ((a+b)(a+c) + (c+d)(b+d))}$$

ARI has the lowest expected value zero for independent clusterings and maximum value 1 for identical clusterings. At the same time, we minimize the equivalence mismatch distance, termed Mirkin Index, that counts all disagreements pairs *b* and *c* as follows:

$$\mu_2 = (b+c)/(a+b+c+d)$$

Received: 1 June 2015 Accepted: 1 April 2016

Published online: 26 April 2016

### References

- Eitel C, Hindricks G, Dagres N, Sommer P, Piorkowski C. Ensite velocity cardiac mapping system: a new platform for 3d mapping of cardiac arrhythmias. *Expert Rev Med Devices*. 2010;7(2):185–92.
- Calkins H, Kuck KH, Cappato R, Brugada J, Camm AJ, Chen S-A, Crijns HJG, Jr Damiano RJ. 2012 hrs/ehra/ecas expert consensus statement on catheter and surgical ablation of atrial fibrillation. *Heart Rhythm*. 2012;9(4):632–696.e21.
- Konings KTS, Smeets JLRM, Penn OC, Wellens HJJ, Allessie MA. Configuration of unipolar atrial electrograms during electrically induced atrial fibrillation in humans. *Circulation*. 1997;95:1231–41.
- Zlochiver S, Yamazaki M, Kalifa J, Berenfeld O. Rotor meandering contributes to irregularity in electrograms during atrial fibrillation. *Heart R*. 2008;5(6):846–54.
- Chang S-L, Chen Y-C, Hsu C-P, Kao Y-H, Lin Y-K, Lin Y-J, Wu T-J, Chen S-A, Chen Y-J. Electrophysiological characteristics of complex fractionated electrograms and high frequency activity in atrial fibrillation. *Int J Cardiol*. 2013;168(3):2289–99.
- Nademanee K. Trials and travails of electrogram-guide ablation of chronic atrial fibrillation. *Circulation*. 2007;115(20):2592–4.
- Nademanee K, McKenzie J, Kosar E, Schwab M, Sunsaneewitayakul B, Vasavakul T, Khunnawat C, Ngarmukos T. A new approach for catheter ablation of atrial fibrillation: mapping of the electrophysiologic substrate. *J Am Coll Cardiol*. 2004;43(11):2044–53.
- Berenfeld O, Jalife J. Complex fractionated atrial electrograms: is this the beast to tame in atrial fibrillation? *Circ Arrhythm Electrophysiol*. 2011;4(4):426–8.
- Orlov MV. A farewell to arms: Are complex fractionated atrial electrograms doomed as a target for af ablation? *Heart Rhythm*. 2011;8:1720–1.
- Nollo G, Marconcini M, Faes L, Bovolo F, Ravelli F, Bruzzese L. An automatic system for the analysis and classification of human atrial fibrillation patterns from intracardiac electrograms. *IEEE Trans Biomed Eng*. 2008;55(9):2275–85.
- Hunter RJ, Diab I, Thomas G, Duncan E, Abrams D, Dhinoja M, Sporton S, Earley MJ, Schilling RJ. Validation of a classification system to grade fractionation in atrial fibrillation and correlation with automated detection systems. *Europace*. 2009;11(12):1587–96.
- Scherr D, Dalal D, Cheema A, Cheng A, Henrikson CA, Spragg D, Marine JE, Berger RD, Calkins H, Dong J. Automated detection and characterization of complex fractionated atrial electrograms in human left atrium during atrial fibrillation. *Heart Rhythm*. 2007;4(8):1013–20.
- Faes L, Nollo G, Antolini R, Gaita F, Ravelli F. A method for quantifying atrial fibrillation organization based on wave-morphology similarity. *IEEE Trans Biomed Eng*. 2002;49(12):1504–13.
- Sanders P, Berenfeld O, Hocini M, Jais P, Vaidyanathan R, Hsu L-F, Garrigue S, Takahashi Y, Rotter M, Sacher F, Scavee C, Ploutz-Snyder R, Jalife J, Haissaguerre M. Spectral analysis identifies sites of high-frequency activity maintaining atrial fibrillation in humans. *Circulation*. 2005;112(6):789–97.

15. Ganesan AN, Kuklik P, Lau DH, Brooks AG, Baumert M, Lim WW, Thanigaimani S, Nayyar S, Mahajan R, Kalman JM, Roberts-Thomson KC, Sanders P. Bipolar electrogram shannon entropy at sites of rotational activation: implications for ablation of atrial fibrillation. *Circ Arrhythm Electrophysiol*. 2013;6:48–57.
16. Ugarte JP, Orozco-Duque A, Tobón C, Kremen V, Novak D, Saiz J, Oesterlein T, Schmitt C, Luik A, Bustamante J. Dynamic approximate entropy electroanatomic maps detect rotors in a simulated atrial fibrillation model. *Plos One*. 2014;9:e114577.
17. Lau DH, Zeemering S, Maesen B, Kuklik P, Verheule S. Catheter ablation targeting complex fractionated atrial electrogram in atrial fibrillation. *J Atr Fibrillation*. 2013;6(3):24–6.
18. Ravelli F, Mase M. Computational mapping in atrial fibrillation: how the integration of signal-derived maps may guide the localization of critical sources. *Europace*. 2014;16(5):714–23.
19. Barbaro V, Bartolini P, Calcagnini G, Morelli S, Michelucci AGG. Automated classification of human atrial fibrillation from intraatrial electrograms. *Pacing Clin Electrophysiol*. 2000;23(2):192–202.
20. Schilling C, Keller M, Scherr D, Oesterlein T, Haissaguerressaguerre M, Schmitt C, Dossel O, Luik A. Fuzzy decision tree to classify complex fractionated atrial electrograms. *Biomed Tech (Berl)*. 2015;60(3):245–55.
21. Almeida TP, Chu G, Salinet JL, Schlindwein FS. Minimizing discordances in automated classification of fractionated electrograms in human persistent atrial fibrillation. *Med Biol Eng Comput*. 2015.
22. Porter M, Spear W, Akar J, Helms R, Brysiewicz N, Santucci P, Wilber D. Prospective study of atrial fibrillation termination during ablation guided by automated detection of fractionated electrograms. *J Cardiovasc Electrophysiol*. 2008;19(6):613–20.
23. Singh JP, Ptaszek LM, Verma A. Elusive atrial substrate: Complex fractionated atrial electrograms and beyond. *Heart Rhythm*. 2010;7(12):1886–90.
24. Botteron GW, Smith JM. 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*. 1995;42(6):579–86.
25. Chen W, Zhuang J, Yu W, Wang Z. Measuring complexity using fuzzyen, apen, and sampen. *Med Eng Phys*. 2009;31(1):61–8.
26. Orozco-Duque A. Multifractal analysis for grading complex fractionated electrograms in atrial fibrillation. *Physiol Meas*. 2015;36(11):2269–84.
27. Filippone M, Camastra F, Masulli F, Rovetta S. A survey of kernel and spectral methods for clustering. *Pattern Recognit*. 2008;41:176–90.
28. Nascimento M, Carvalho A. Spectral methods for graph clustering - a survey. *Eur J Oper Res*. 2011;211:221–31.
29. Santos J, Embrechts M. On the use of the adjusted rand index as a metric for evaluating supervised classification. *Lect Notes Comput Sci: Artifi Neural Netw - ICANN*. 2009;2009(5769):175–84.
30. Alvarez-Meza AM, Cardenas-Pena D, Castellanos-Dominguez G. Unsupervised kernel function building using maximization of information potential variability. *Lect Notes Comput Sci: Prog Pattern Recognit, Image Anal, Comput Vis, Appl*. 2014;8827:335–42.
31. Mehrkanoon S, Alzate C, Mall R, Langone R, Suykens JA. Multi-class semi-supervised learning based upon kernel spectral clustering. *IEEE Trans Neural Netw Learn Syst*. 2015;26(4):720–33.
32. Kremen V. Automated assessment of endocardial electrograms fractionation in human. PhD thesis, The Czech Technical University in Prague. 2008.
33. Ravelli F, Mase M, Cristoforetti A, Marini M, Disertori M. The logical operator map identifies novel candidate markers for critical sites in patients with atrial fibrillation. *Prog Biophys Mol Biol*. 2014;115:186–97.
34. Navoret N, Jacquir S, Laurent G, Binczak S. Detection of complex fractionated atrial electrograms using recurrence quantification analysis. *IEEE Trans Biomed Eng*. 2013;60(7):1975–82.
35. Lin Y-J, Lo M-T, Lin C, Chang S-L, Lo L-W, Hu Y-F, Hsieh W-H, Chang H-Y, Lin W-Y, Chung F-P, Liao J-N, Chen Y-Y, Hanafy D, Huang NE, Chen S-A. Prevalence, characteristics, mapping, and catheter ablation of potential rotors in nonparoxysmal atrial fibrillation. *Circ Arrhythm Electrophysiol*. 2013;6(5):851–8.
36. Hunter RJ, Diab I, Tayebjee M, Richmond L, Sporton S, Earley MJ, Schilling RJ. Characterization of fractionated atrial electrograms critical for maintenance of atrial fibrillation: a randomized, controlled trial of ablation strategies (the cfae af trial). *Circ Arrhythm Electrophysiol*. 2011;4(5):622–9.
37. Kalifa J, Tanaka K, Zaitsev AV, Warren M, Vaidyanathan R, Auerbach D, Pandit S, Vikstrom KL, Ploutz-Snyder R, Talkachou A, Aتيenza F, Guiraudon G, Jalife J, Berenfeld O. Mechanisms of wave fractionation at boundaries of high-frequency excitation in the posterior left atrium of the isolated sheep heart during atrial fibrillation. *Circulation*. 2006;113(5):626–33.
38. Wu SH, Jiang WF, Gu J, Zhao L, Wang YL, Liu YG, Zhou L, Gu JN, Xu K, Liu X. Benefits and risks of additional ablation of complex fractionated atrial electrograms for patients with atrial fibrillation: A systematic review and meta-analysis. *Int J Cardiol*. 2013;169(1):35–43.