Supplementary Material for:

Feature subset selection and classification of intracardiac electrograms during atrial fibrillation.

S.I Duque, A Orozco-Duque, V Kremen, D Novak, C Tobón and J Bustamante

EGM signal characterization – feature extraction methods

This supplementary material presents detailed explanation about features extraction methods used in this study:

Local Activation Waves Detection

The Local Activation Wave (LAW) was carried out as follow: i) The continuous wavelet transform (CWT) was applied to each EGM. 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 intersegments if the intersegment is shorter than the time window (40 ms). iv) We applied an adaptive threshold, based on the algorithm described by Pan and Tompkins [1]. v) EGM segments with amplitude greater than the threshold were detected. Their maximum was marked as LAW_{center}.

Let $x(k), k = 1, \dots, N$ being an EGM with N sample. If the number of LAW_{center} detected by th algorithm is M, we define $y = \{t_{LAWcenter}(1), \dots, t_{LAWcenter}(M)\}$, where $t_{LAWcenter}(i)$ represents the time of the LAW detections. A LAW is a segment of 90 ms [2] and is defined as $LAW_i = x(k), \forall\ k \in [t_{LAWcenter}(i) - 45\ ms, t_{LAWcenter}(i) + 45\ ms]$, $i = 1,2, \dots, M$. Using $y(i)$ and LAW_i the following features were calculated: intervals between LAW $((t_{LAWcenter}(i+1) - t_{LAWcenter}(i))$; number of LAWs detected; and the difference in amplitude between maximum and minimum in each LAW.

We used a zero-crossing detector of the first derivative along the signal. However, to avoid the influence of noise, we used an adaptive threshold th, which is adjusted according to the maximum amplitude as follow, for each LAW_i: $th(i) = th(0) + 0.2 * \max_{LAW_i} x$, where *x* is the signal, and $th(0)$ is the voltage threshold at *t=0*. $th(0)$ is computed as the root mean square value of *x*, and is divided by an introduced sensitivity parameter *a=3*. Using this information, we computed the following features: The number of zero crossings (ZC); the number of max-min intervals; and the ratio between zero crossings and $LAWs$.

Similarity index calculation

The Similarity index proposed in Faes et al.[2] was calculated, using the synchronization process described in section 2.3.2. This index quantifies the regularity of a signal based on the degree of repeatability of the activation waves. We computed two different similarity index ($\rho_1$and ρ_2) as follow:

$$\rho_1 = \frac{2}{M(M-1)} \sum_{i=1}^{M} \sum_{j=i+1}^{M} \theta(\varepsilon - arcos(s_i . s_j))$$

where θ is the Heaviside function, ε is a threshold defined as 0.8, and s are the synchronized and normalized LAW [2].

Based on Pearson's correlation, another similarity index was calculated:

$$\rho_2 = \frac{2}{M(M-1)} \sum_{i=1}^{M} \sum_{j=i+1}^{M} \frac{\sigma_{s_i s_j}}{\sigma_{s_i} \sigma_{s_j}}$$

Where $\sigma_{s_i s_j}$ is the covariance (s_i, s_j), σ_{s_i} is the variance of segment s_i , and σ_{s_j} is the variance of segment s_j

Templates calculation

The template is a general representation of LAW in a given EGM and the baseline of each signal. To calculate the template of the LAW, each LAW_i were synchronized with respect to LAW_1 and was defined as LAW'_i. The template T_LAW represents the morphology of the activation waves for each signal. It was calculated as follows:

$$T\_LAW = \frac{1}{M} \sum_{i=1}^{M} LAW'_i$$

We also calculated a template for the baseline. To this end, we extracted baseline segments taken from the middle point between two activation waves. The middle points were defined as:

$$mp_i = \frac{LAW'_{center\ i+1} - LAW'_{center\ i}}{2} + LAW'_{center\ i}, i = 1,2, \dots, M.$$

The points for the baseline segments were defined as: $bl_i = x(k) , \forall \ k \in [mp_i - 45, mp_i + 45]$. An average baseline template T_bl was calculated from the extracted segments $\{bl_1, bl_2, \dots, bl_{M-1}\}$ for each signal, as follow:

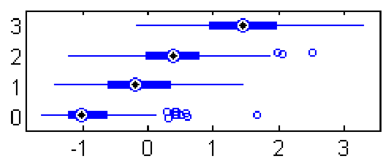
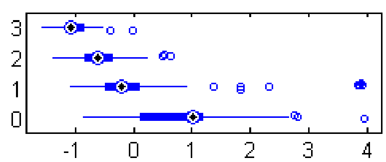
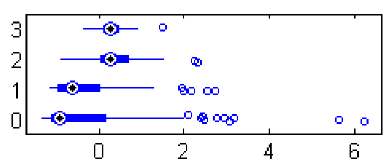
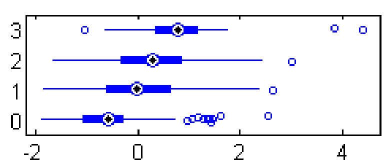
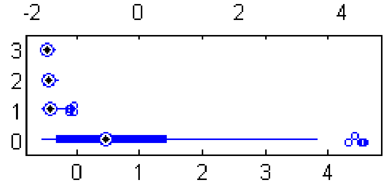
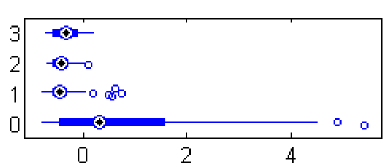
$$T\_BL = \frac{1}{M-1} \sum_{i=1}^{M-1} bl_i$$

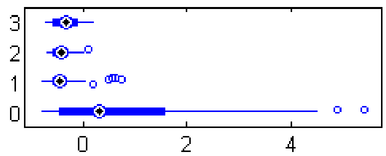
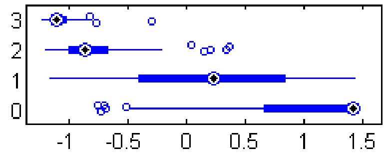
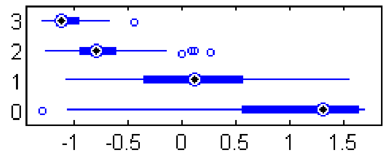
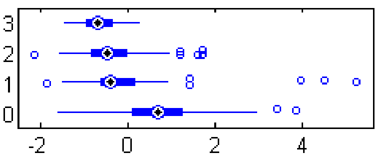
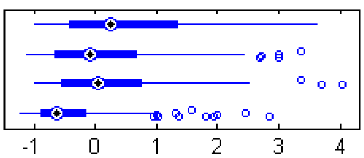
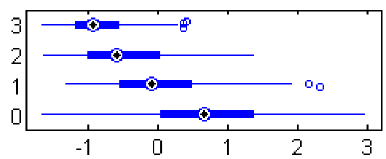
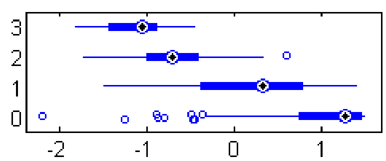
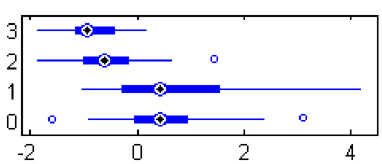
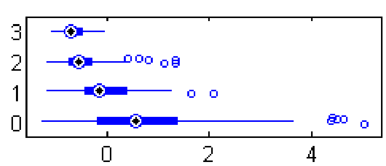
Calculation of segments with continuous local electrical activity

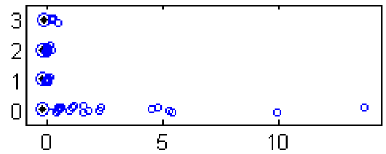
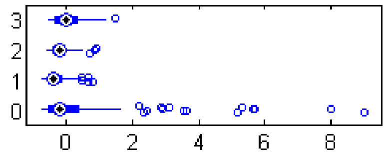
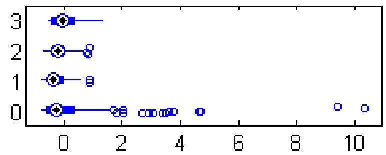
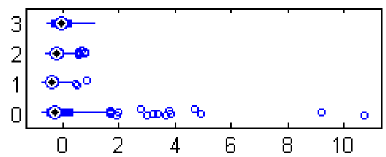
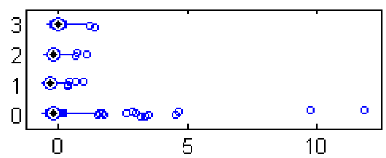
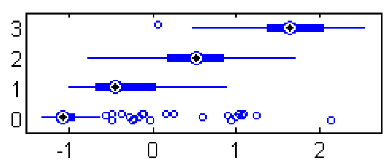
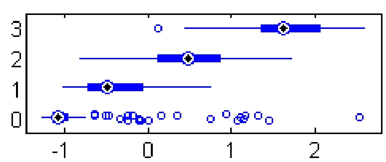
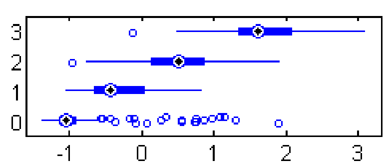
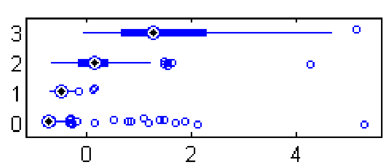
Segments with Continuous Local Electrical Activity (SLEA) were computed for each EGM based on the scheme proposed by Kremen et al.[3]. SLEAs were defined as adjacent primary local activity segments whose intersegment space is less than 40 ms. Primary local activity is detected from the reconstructed signal using scales 1, 2 and 3 in the Continuous Wavelet Transform of the EGM. Two measures were computed: 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 local electrical activity (*LA*). They were calculated as the sum of the segments, where the signal is greater than the threshold before applying intersegmental fusion. The algorithm extracted features based on *SLEAs*, t_{SLEA} and the relation between t_{SLEA} and t_{LA}.

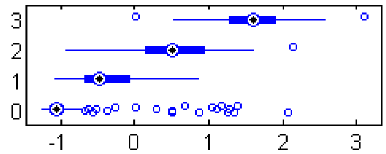
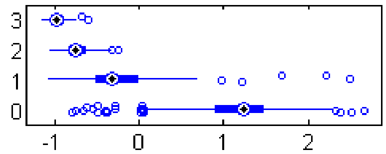
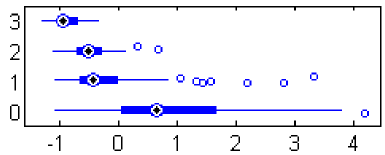
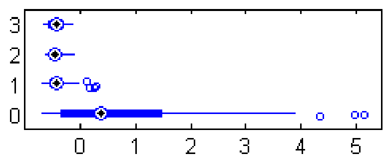
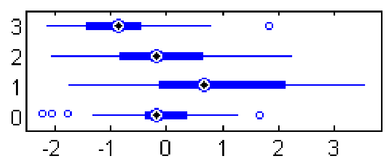
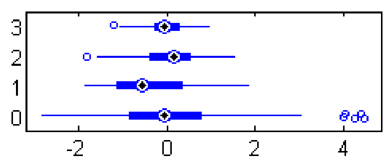
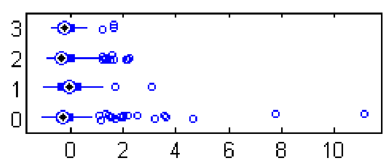
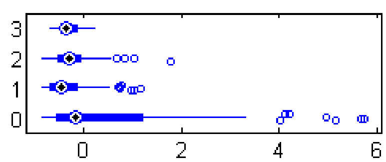
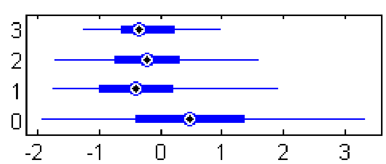
SET OF FEATURES

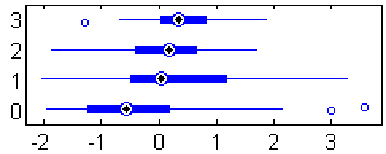
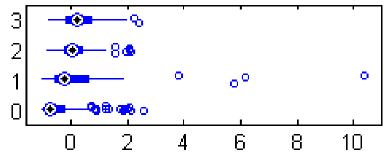
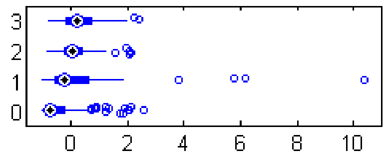
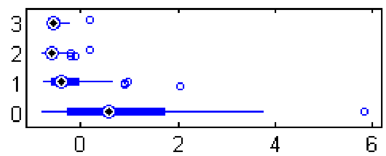
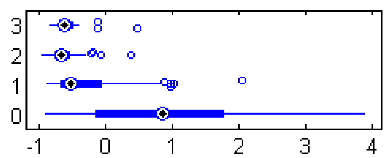
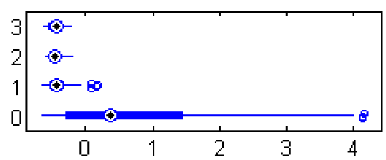
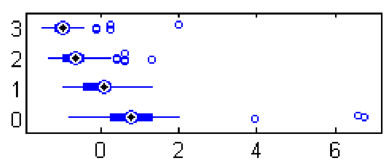
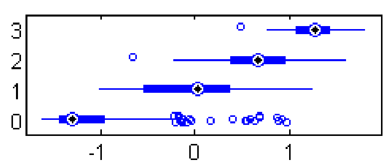
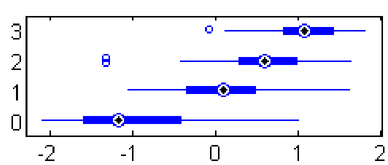
Below is a list of the 42 features obtained for the optimal subset found by genetic algorithms. This list includes the boxplot showing the distribution of the signals in each class for each feature.

Parameter	Description	Boxplot	References
GROUP 1	**Features based on temporal information and on detecting local activation waves**		
$\sum ZC$	Total number of zero crossings in the signal		-
ILAWm	Average of intervals between *LAW*		-
ILAWs	Standard deviation of intervals between *LAW*		-
$\sum ZC/\sum LAW$	Ratio of zero crossings to *LAW*		-
MMm	Average peak-to-peak amplitude in max-min pairs		-
IMMm	Average of intervals between max-min pairs.		-

IMMs	Standard deviation of intervals between max-min pairs.		-
GROUP 2		**Features based on the LAW morphology**	
ρ_1	Similarity index (cosine distance)		[2]
ρ_2	Similarity index (correlation)		[2]
$\frac{1}{M}\sum_1^M \max(LAW_i)$ synchronized	Average of the maximum value of the synchronized *LAW*		-
T_LAWe	Energy of the template of the *LAW*		-
T_LAWcs	Cumulative sum of the template of the *LAW*		-
T_LAWsvd	Singular value decomposition of the template of the *LAW*		[4]
T_LAWam	Average absolute value of the template of the *LAW*		-
EMDT_LAWam	Average absolute value of the empirical mode decomposition of the of the template of the *LAW*		[5]

T_BLe	Energy of the template of the baseline		-
T_BLs	Standard deviation of the template of the baseline		-
T_BLamx	Absolute máximum of the template of the baseline		-
T_BLppmx	Maximum peak-to-peak amplitude of the template of the baseline		-
T_BLcs	Cumulative sum of the template of the baseline		-
GROUP 3	**Features based on segments with continuous local electrical activity (SLEA)**		
$\sum_i t_{SLEA}(i)$	Total value of the width of the *SLEA*		[3]
$\sum_i t_{SLEA}(i) - \sum_i t_{LA}(i)$	Difference between the width of the SLEA and the width of the local electrical activity(t_{LA})		[3]
$\sum SLEA - \sum LA$	Difference between joined segments(SLEA) and separate segments(LA)		[3]
SLEAZCs	Standard deviation of ZC within SLEA		[3]

SLEAZCt	Total of ZC within SLEA		[3]
nSLEAZCm	Average of ZC without SLEA		[3]
nSLEAZCs	Standard deviation of ZC without SLEA		[3]
GROUP 4	**Features based on the frequency domain and the time-frequency domain**		
PSmf	Mean frequency in the power spectrum		-
RI	Regularity index		-
PSs	Standard deviation of the power spectrum		-
KC	Coefficient of kurtosis		-
EDWTd2	Energy in detail coefficient 2 in DWT (Discrete Wavelet Transform, Daubechies 6, 8 levels)		-
EDWTd3	Energy in detail coefficient 3 in DWT		-

EDWTd5	Energy in detail coefficient 5 in DWT		-
EDWTd6	Energy in detail coefficient 6 in DWT		-
EDWTd8	Energy in detail coefficient 8 in DWT		-
GROUP 5	**Features based on information theory and non-linear dynamics**		
ACORRe	Energy of the autocorrelation signal		-
ACORRs	Standard deviation of the autocorrelation signal		-
SVDx	Singular value decomposition of the signal		[4]
ACORRp	Number of significant peaks in the autocorrelation signal *(using a threshold th=0.2*RMS(x))*		-
ApEn	Approximate Entropy		[6]
MI	Mutual information between subsegments of the signal (300 ms)		[7]

Below is a list of the 50 remaining features which were removed from the optimal subset:

GROUP 1	GROUP 2
• $\sum ZC$ in x (threshold) • Average of intervals between LAW (uncorrected) • $\sum LAW$ in x • Average of intervals between ZC • Standard deviation of intervals between ZC • Standard deviation peak-to-peak amplitude in max-min pairs • Average of the peak-to-peak amplitude in LAW • Standard deviation of the peak-to-peak amplitude in LAW • Total Inflexions	• ZC in synchronized LAW • Average of the peak-to-peak amplitude in synchronized LAW • Standard deviation of the peak-to-peak amplitude in synchronized LAW • RMS of T_LAW • Standard deviation of T_LAW. • Energy of continuous wavelet coefficients of T_LAW • Absolute maximum of T_LAW • Peak-to-peak amplitude of T_LAW • Energy of the Empirical mode decomposition of T_LAW • RMS of the Empirical mode decomposition of T_LAW • Standard deviation of the Empirical mode decomposition of T_LAW • ZC in the Empirical mode decomposition of T_LAW • Absolute maximum of the Empirical mode decomposition of T_LAW • Peak-to-peak amplitude of the Empirical mode decomposition of T_LAW • Cumulative sum of the Empirical mode decomposition of T_LAW • Singular value decomposition of the Empirical mode decomposition of T_LAW • RMS of T_bl • Standard deviation T_bl • ZC in T_bl • Singular value decomposition of T_bl • Average absolute value of T_bl
GROUP 3 [3] • $\frac{1}{M}\sum_{1}^{M} t_{SLEA}(i)$ • Standard deviation of t_{SLEA} • $\sum SLEA \,/ \sum LA$ • Average of ZC within SLEA • $\sum ZC$ without SLEA	**GROUP 4** • Maximum in the power spectrum • Dominant frequency in the power spectrum • Peak frequency in the power spectrum • First quartile frequency in the power spectrum • Third quartile frequency in the power

	spectrum • Interquartile range of the power spectrum • Frequency with 95% of energy in the power spectrum • Coefficient of asymmetry • Total energy of the power spectrum • Energy in approximation coefficient in DWT • Energy in detail coefficient 1 in DWT • Energy in detail coefficient 4 in DWT • Energy in detail coefficient 7 in DWT
GROUP 5 • Peaks of the autocorrelation signal • Shannon Entropy [8]	

References

[1] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm.," *IEEE Trans. Biomed. Eng.*, vol. 32, no. 3, pp. 230–6, Mar. 1985.

[2] L. Faes, G. Nollo, R. Antolini, F. Gaita, and F. Ravelli, "A method for quantifying atrial fibrillation organization based on wave-morphology similarity.," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 12 Pt 2, pp. 1504–13, Dec. 2002.

[3] V. Kremen, L. Lhotská, M. Macas, R. Cihák, V. Vancura, J. Kautzner, and D. Wichterle, "A new approach to automated assessment of fractionation of endocardial electrograms during atrial fibrillation.," *Physiol. Meas.*, vol. 29, no. 12, pp. 1371–81, Dec. 2008.

[4] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numer. Math.*, vol. 14, no. 5, pp. 403–420, 1970.

[5] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 1998, vol. 454, no. 1971, pp. 903–995.

[6] S. M. Pincus, "Approximate entropy as a measure of system complexity.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 88, no. 6, pp. 2297–301, Mar. 1991.

[7] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A*, vol. 33, no. 2, p. 1134, 1986.

[8] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Inf. theory*, vol. 37, no. 1, pp. 145–151, 1991.