Unsupervised classification of cardiac arrhythmias using partitional clustering

D. Peluffo, J. L. Rodríguez, G. Castellanos

ABSTRACT

This work describes a method to classify cardiac arrhythmias in Holter records using unsupervised classification. Unsupervised analysis is preferred in this work because detecting a specific heartbeat in Holter records requires analyzing every heartbeat taking into account several factors such as: variability, signal length, EMG noise, artifacts and different dynamic behavior and morphology. ECG signals are extracted from MIT-BIH arrhythmia database. The proposed method includes pre-processing, feature selection and unsupervised classification. Features are obtained from heartbeat morphology, time related and representation features. For feature selection, we apply the unsupervised $Q - \alpha$ algorithm obtained from spectral and graph-based analysis. The classification stage is based on partitional clustering using the general iterative model. In order to improve the algorithm convergence, clustering incorporates a center initialization based on J-means criterion. The method shows good average performance, namely sensitivity 98.9%, specificity 99.8 % and clustering performance 99.6%.

Key words **- Cardiac arrhythmias, clustering, Holter records, spectral analysis**

I. INTRODUCTION

For outpatient electrocardiography there exist the Holter records which are recorded for long time and allow to assess the heart condition without altering the patient daily activities. Thus, they are useful to detect pathologies that are hard to diagnose in short-time ECG (12 leads). The problem of these records is the wide amount of heartbeats which complicates its visual inspection. For this reason, computer analysis systems have been developed and are commonly used as diagnostic support.

In general, these systems work off-line and must take into account some factors that add variability, such as: signal length, artifacts, EMG noise and different dynamic behavior and morphology (different patient and/or pathology). Therefore, it is necessary to analyze each heartbeat, in detail, to detect a specific heartbeat. For this reason, unsupervised classification is preferred in this approach. Clustering is the most frequently used technique for unsupervised analysis of ECG signals.

The aim of this work is to propose a system for unsupervised classification of ECG signals using partitional clustering. The proposed system includes: pre-processing, feature selection and unsupervised classification. The experimental set used was drawn from the MIT-BIH database [1], including normal (N) heartbeats, as well as the arrhythmia types recommended by the AAMI [2], usually found in Holter registers: ventricular extrasystoles (V), left (L) and right (R) branch bundle blocks and atrial premature beats (A). The set of features are ranked by means of a non-supervised scheme based on partitional clustering algorithms such as H-means and density based clustering. These algorithms were implemented using a general iterative model for center-based clustering [3]. Center initialization is carried out using J-means algorithm [4] and its variants. J-means variants consist of some changes in the objective function. The results were assessed in terms of sensitivity and specificity measures, based on the heartbeat labels. Additionally, we take into account another performance measures, namely sensitivity to initialization, convergence time and a non-supervised index for clustering.

II. METHODOLOGY

The proposed system includes pre-processing, feature selection and unsupervised classification stages. The following explains each stage in the system.

Diego Peluffo-Ordóñez: dhpeluffoo@unal.edu.co, master student, Universidad Nacional de Colombia - Manizales.

José Luis Rodríguez-Sotelo: jlrodriguezso@unal.edu.co, PhD student, Universidad Nacional de Colombia - Manizales.

Germán Castellanos-Domínguez: cgcastellanosd@unal.edu.co, PhD, Profesor, Department of electrical engineering, electronic engineering and computer sciences of Universidad Nacional de Colombia - Manizales.

This work was supported by GC&PDS research group.

A. Pre-processing

Signals (heartbeats, QRS complexes and P waves) were normalized using:

$$\boldsymbol{y}_n = \frac{\boldsymbol{y} - \mu(\boldsymbol{y})}{|\boldsymbol{y}|_{max}} \tag{1}$$

where $\mu(\cdot)$ represents the mean.

B. Feature set

The data set are features taken from previous works that achieved good performance in characterization of wave morphology, signal variability, and signal representation. They have been employed in applications to detect N, V [5], L [6], R [6], [7] and A [8]. Table I shows, in detail, the feature set, including its description and computation. The segmentation stages were implemented as is described in [6] and [8], [5].

C. Feature selection

[10] provides a definition of relevance in terms of an *affinity* matrix which captures the inner products of the observations and a weighting vector. This concept is based on consistency of resulting subsets or clusters from a clustering process using spectral properties and topological analysis (graph theory), where the data matrix represents the vertices of an undirected graph and the affinity matrix indicates the weights of each edge [11]. This feature selection method is called $Q-\alpha$ because $\boldsymbol{Q}$ is the orthonormal rotation matrix and $\boldsymbol{\alpha}$ represents the weighting vector. In this study, the orthonormal matrix is denoted by $\boldsymbol{U}$.

Let $\boldsymbol{M}$ be a $q \times n$ matrix: $\boldsymbol{M} = \boldsymbol{X}^\mathsf{T} = (\boldsymbol{m}_1, \ldots, \boldsymbol{m}_q)^\mathsf{T}$ which satisfies $\|\boldsymbol{m}_i\| = 1$ and $\mu(\boldsymbol{m}_i) = 0$. Given this, the affinity matrix can be computed as:

$$\boldsymbol{C}_{\boldsymbol{\alpha}} = \sum_{i=1}^{q} \alpha_i \boldsymbol{m}_i \boldsymbol{m}_i^\mathsf{T} = \boldsymbol{M}^\mathsf{T} \ \mathrm{Diag}(\boldsymbol{\alpha}) \boldsymbol{M} \tag{2}$$

where $\mathrm{Diag}(\cdot)$ denotes a diagonal matrix formed by its argument vector.

From a statistical point of view, the general idea of the $Q-\alpha$ algorithm is the same as PCA, where the accumulated variability is measured by means of $\boldsymbol{C}_{\boldsymbol{\alpha}}$ instead of the covariance matrix. The scaling factor α_i allows to adjust the relative importance of each observation. Given that the first orthonormal matrix is chosen randomly, the selection of relevant features is directly related to the value of α_i. Then, whereas the solution of the optimization problem is to find the best vector $\boldsymbol{\alpha}$, we can formulate the following optimization problem:

$$\max_{\boldsymbol{\alpha}} \ \mathrm{tr}(\boldsymbol{U}^\mathsf{T} \boldsymbol{C}_{\boldsymbol{\alpha}} \boldsymbol{C}_{\boldsymbol{\alpha}} \boldsymbol{U}) = \sum_{j=1}^{p} \lambda_j^2 \tag{3}$$

$$\text{s. t.} \quad \boldsymbol{U}^\mathsf{T}\mathbf{U} = \boldsymbol{I}_p, \quad \boldsymbol{\alpha}^\mathsf{T}\boldsymbol{\alpha} = 1 \tag{4}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ are the eigenvalues of $\boldsymbol{C}_{\boldsymbol{\alpha}}$ and $\boldsymbol{U}$ is a random orthonormal matrix.

To duplicate the term of affinity matrix does not affect the objective function because $\boldsymbol{C}_{\boldsymbol{\alpha}}$ is a symmetric and semi-positive definite matrix, furthermore, in mathematical terms, this is advantageous because it allows us to suggest a quadratic form with respect to the variable of interest, so:

$$\max_{\boldsymbol{\alpha}} \boldsymbol{\alpha}^\mathsf{T} \boldsymbol{G} \boldsymbol{\alpha} \tag{5}$$

$$\text{s. t.} \quad \boldsymbol{\alpha}^\mathsf{T}\boldsymbol{\alpha} = 1 \tag{6}$$

where $\boldsymbol{G}$ is an auxiliar matrix whose elements are $G_{ij} = (\boldsymbol{m}_i^\mathsf{T} \boldsymbol{m}_j) \boldsymbol{m}_i^\mathsf{T} \boldsymbol{U} \boldsymbol{U}^\mathsf{T} \boldsymbol{m}_j$.

The previous equations correspond to the objective function of the unsupervised version of $Q-\alpha$. Given that the matrix $\boldsymbol{G}$ is obtained from an arbitrary orthonormal transformation, it is necessary to apply an iterative method which tunes $\boldsymbol{U}$ and $\boldsymbol{\alpha}$. From the optimization problem (3), we can note that the vector $\boldsymbol{\alpha}$ indicates the direction of the relevant features and the matrix $\boldsymbol{U}$ means its rotation, and therefore the adjustment of these parameters is mutually dependent and must be achieved on an alternant way, as shown in Algorithm 1.

Algorithm 1 Power-embedded $Q-\alpha$ method

1. Initialization: $\boldsymbol{M} = \boldsymbol{X}^\mathsf{T}$, chose at random $k \times n$ matrix $\boldsymbol{U}^{(0)}$ $(\boldsymbol{U}^{(0)\mathsf{T}} \boldsymbol{U}^{(0)} = \boldsymbol{I}_n)$, $\boldsymbol{m}_i \leftarrow (\boldsymbol{m}_i - \mu(\boldsymbol{m}_i))/\|\boldsymbol{m}_i\|$.
2. Form $\boldsymbol{G}$: $G_{ij} = (\boldsymbol{m}_i^\mathsf{T} \boldsymbol{m}_j) \boldsymbol{m}_i^\mathsf{T} \boldsymbol{U} \boldsymbol{U}^\mathsf{T} \boldsymbol{m}_j$
3. Compute $\boldsymbol{\alpha}$ as the eigenvector related to the major eigenvalue of $\boldsymbol{G}$.
4. Compute $\boldsymbol{C}_{\boldsymbol{\alpha}} = \boldsymbol{M}^\mathsf{T} \ \mathrm{Diag}(\boldsymbol{\alpha}) \boldsymbol{M}$
5. Obtain the orthonormal transformation: $\boldsymbol{Z}^{(r)} = \boldsymbol{C}_{\boldsymbol{\alpha}}^{(r)} \boldsymbol{U}^{(r-1)}$
6. QR decomposition: $[\boldsymbol{U}^{(r)}, \boldsymbol{R}] = \mathrm{qr}\,(\boldsymbol{Z}^{(r)})$
7. $r \leftarrow r + 1$ and return to the step 2

This method achieves higher performance and is significantly more efficient than the basic method described in [10], which from an initial vector $\boldsymbol{\alpha}$

Feature	Type	Description
1	HRV [8]	• RR interval (r_1)
2		• pre-RR interval (r_0)
3		• post-RR interval (r_2)
4	Prematurity [8]	• RR and pre-RR interval difference, $f_4 = r_1 - r_0$
5		• post-RR and RR interval difference, $f_5 = r_2 - r_1$
6		• continuous heartbeats type A $\circ$ $f_6 = \left(\frac{r2}{r1}\right)^2 + \left(\frac{r0}{r1}\right)^2 - \left(f_4 + \beta \cdot \frac{1}{3}\sum_{i=0}^{2} r_i^2 \cdot \log(r_i)^2\right)$, $0 < \beta < 0.1$
7	Morphology [8], [5]	• Dynamic time warping (DTW) between current P wave and average P wave
8		• QRS complex polarity $\circ$ Let b be the samples of a heartbeat, then $f_8 = P_i = \left\lvert \frac{\max(b_i)}{\min(b_i)} \right\rvert$
9		• QRS complex energy $\circ$ $f_9 = E_{b_j} = \sum_{i=1}^{n} b_j(i)^2$
$10, \ldots, 19$	Representation [7]	• 10 Hermite coefficients $f_{\substack{i \\ i=10:19}} = C_n^\sigma = y.\phi_{\substack{n \\ n=0:9}}^\sigma$, where $\phi_{\substack{n \\ n=0:9}}^\sigma = \frac{e^{-t^2/2^{\sigma^2}}}{\sqrt{2^n n! \sqrt{\pi}}} H_n(t/\sigma)$, H_n is a Hermite polynomial, y is the qrs complex and σ is the window width
$20, \ldots, 90$	Representation [9]	• Wavelet Db2 (A4: $20-25$, D4: $26-31$, D3: $32-41$, D2: $43-58$, D1: $59-90$) $\circ$ Using Wavelet Discrete Transform (DWT)

TABLE I
FEATURE SET USED IN ARRHYTHMIA ANALYSIS

computes $\boldsymbol{Q}$ iteratively (and viceversa) [12].

[10] also presents two alternatives of unsupervised $Q - \alpha$: by using a normalization through Laplacian matrix $\boldsymbol{C_\alpha}$, and another one based on Ritz acceleration and singular value decomposition when one assumes a initial vector $\boldsymbol{\alpha}$. Additionally, they demonstrate the convergence of the algorithm. Nevertheless, an indicator of the algorithm convergence could be the change of the vector $\boldsymbol{\alpha}$, i.e, the difference between the current and preceding vector: $\|\boldsymbol{\alpha}^{(r)} - \boldsymbol{\alpha}^{(r-1)}\| < \delta$, where δ defines accuracy.

In steps 5 and 6, an orthonormal projection of $\boldsymbol{C_\alpha}$ and QR decomposition are applied, respectively, to obtain the refined matrix $\boldsymbol{U}$ for the next iteration.

Finally, the p relevant features correspond to the elements of $\boldsymbol{M}$ that satisfy $\sum_{i=1}^{p} \alpha_i^2 \approx N/100$, for $N\%$ of accumulated variance.

D. Unsupervised classification

The classical technique of unsupervised classification (grouping) is the partitional clustering or center-based clustering (CBC) which has the goal of minimizing an objective function to obtain an optimal solution via iterative updating-centers [3]. The objective function defines how good a clustering solution is and must be related to the updating-centers function [13].

For example, in the minimum sum of squares based clustering (*MSSC*), the objective function is:

$$\min_{\boldsymbol{\rho}_k \in \boldsymbol{P}_k} \sum_{j=1}^{k} \sum_{\boldsymbol{x}_l \in \boldsymbol{C}_l} ||\boldsymbol{x}_l - \boldsymbol{q}_j||^2 \tag{7}$$

where $\|\cdot\|$ represents the Euclidean norm, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\mathsf{T}$ is the data set to be clustered, $\boldsymbol{P}_k$ is the partition set of $\boldsymbol{X}$, k is the number of clusters, $\boldsymbol{C} = \{\boldsymbol{C}_1, \ldots, \boldsymbol{C}_k\}$ is the k-dimensional clusters set and $\boldsymbol{q}_j$ is the center or centroid given by:

$$\boldsymbol{q}_j = \bar{\boldsymbol{x}}_j = \frac{1}{|\boldsymbol{C}_j|} \sum_{\boldsymbol{x}_i \in \boldsymbol{C}_j} \boldsymbol{x}_i \tag{8}$$

where $|\cdot|$ represents the number of points or elements of its argument.

Center-based clustering: The general iterative clustering (GIC) is based on the H-means algorithm. H-means works as follows. An initial partition $\boldsymbol{C} = \{\boldsymbol{C}_1, \ldots, \boldsymbol{C}_k\}$ is chosen at random and the centroids $\boldsymbol{Q} = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k\}$ of each cluster are computed. Then each data point is assigned (reallocated) to its closest centroid $\boldsymbol{q}_j$, if no change in assignments occurs, the heuristic stops with a locally minimum partition. Otherwise, the centroids are updated and the procedure iterated [4].

Define a d-dimensional set of n data points $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\mathsf{T}$ as the data to be clustered and a d-dimensional set of k centroids $\boldsymbol{Q} = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k\}$ as the clustering solution that an iterative algorithm refines. In the general model for clustering algorithms that use iterative optimization, the centroids are computed using a membership function $m(\boldsymbol{q}_j/\boldsymbol{x}_i)$ and a weight function $w(\boldsymbol{x}_i)$, which define the proportion of data point $\boldsymbol{x}_i$ that belongs to center $\boldsymbol{q}_j$ and how much influence data point $\boldsymbol{x}_i$ has in recomputing the centroid parameters for the next iteration, respectively. The membership function satisfies $m(\boldsymbol{q}_j/\boldsymbol{x}_i) \geq 0$ and $\sum_{j=1}^{k} m(\boldsymbol{q}_j/\boldsymbol{x}_i) = 1$. This function is called *hard* when $m \in \{0, 1\}$ and *soft* when $0 \leq m \leq 1$.

Therefore, the steps that describe the GIC for center-based clustering are:

1) Define the initials centers.
2) For each data point $\boldsymbol{x}_i$, compute its m and w values.
3) For each centroid $\boldsymbol{q}_j$ recompute its allocation from all data points $\boldsymbol{x}_i$ using:

$$\boldsymbol{q}_j = \frac{\sum_{i=1}^{n} m(\boldsymbol{q}_j/\boldsymbol{x}_i) w(\boldsymbol{x}_i) \boldsymbol{x}_i}{\sum_{i=1}^{n} m(\boldsymbol{q}_j/\boldsymbol{x}_i) w(\boldsymbol{x}_i)} \tag{9}$$

4) Repeat the steps 2 and 3 until a number of iterations N_{iter} or convergence.

H-means: The objective function that the H-means algorithm optimizes is:

$$HM(\boldsymbol{X}, \boldsymbol{Q}) = \sum_{i=1}^{n} \min_{j \in 1,\ldots,k} ||\boldsymbol{x}_i - \boldsymbol{q}_j||^2 \tag{10}$$

H-means has a hard membership function and fixed weights, as is shown in the following expressions:

$$m_{HM}(\boldsymbol{q}_j/\boldsymbol{x}_i) = \begin{cases} 1 & \text{if } l = \arg\min_j ||\boldsymbol{x}_i - \boldsymbol{q}_j||^2 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

and

$$w_{HM}(\boldsymbol{x}_i) = 1 \tag{12}$$

Note that the objective function of H-means is the same as MSSC. By replacing the expressions (11) and (12) in (9), we obtain the same updating-centers function as MSSC (8). Therefore, the general iterative model is demonstrated.

Density based-clustering: There exist several alternatives to the H-means algorithm using the GIC model. In this work, the parametric and non-parametric density based clustering (DBC) are used: Gaussian expectation-maximization clustering (GEMC) and non-parametric DBC using Parzen's method. These algorithms employ a soft membership function and fixed weights. The GEMC objective function is a linear combination of gaussian distributions centered at each centroid and the goal is maximizing its value [13]. The objective function of GEMC can be written as:

$$GEMC(\boldsymbol{X}, \boldsymbol{Q}) = -\sum_{i=1}^{n} \log \left(\sum_{j=1}^{k} p(\boldsymbol{x}_i/\boldsymbol{q}_j) p(\boldsymbol{q}_j) \right) \tag{13}$$

where $p(\boldsymbol{x}_i/\boldsymbol{q}_j)$ is the probability of $\boldsymbol{x}_i$, since it is generated by a Gaussian distribution centered at $\boldsymbol{q}_j$, and $p(\boldsymbol{q}_j)$ is the prior probability of the cluster whose centroid is $\boldsymbol{q}_j$. The log function is used for simplicity, and the minus sign accounts for minimization. The membership and weight functions for GEMC are:

$$m_{GEMC}(\boldsymbol{q}_j/\boldsymbol{x}_i) = \frac{p(\boldsymbol{x}_i/\boldsymbol{q}_j) p(\boldsymbol{q}_j)}{p(\boldsymbol{x}_i)} \tag{14}$$

and

$$w_{GEMC}(\boldsymbol{x}_i) = 1 \tag{15}$$

The Bayes rule is used to compute m_{GEMC}, where $p(\boldsymbol{x}_i)$ is the evidence defined as $p(\boldsymbol{x}_i) = \sum_{j=1}^{k} p(\boldsymbol{x}_i/\boldsymbol{q}_j)$. In the parametric case, the term $p(\boldsymbol{x}_i/\boldsymbol{q}_j)$ can be computed as a normal gaussian centered at centroid $\boldsymbol{q}_j$ and covariance $\boldsymbol{\Sigma}_j$, i.e., $N(\boldsymbol{q}_j, \boldsymbol{\Sigma}_j)$, where $\boldsymbol{\Sigma}_j$ is the covariance matrix which is computed for each cluster $\boldsymbol{\Sigma}_j = cov(\boldsymbol{C}_j)$ and updating for each iteration.

Parzen's method is used to estimate the membership function in the non-parametric case. For non-parametric density-based clustering (NDBC) using Parzen's estimation, the membership function is the same as GEMC, except that the term $p(\boldsymbol{x}_i/\boldsymbol{q}_j)$ is computed as follows:

$$p(\boldsymbol{x}_i/\boldsymbol{q}_j) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right) \tag{16}$$

where K is the Gaussian kernel which is given by:

$$K(\boldsymbol{z}) = \frac{1}{(2\pi)^{-d/2}} e^{-\frac{1}{2}\boldsymbol{z}\boldsymbol{z}^{\intercal}} \tag{17}$$

E. Center initialization

One of the biggest problems of the clustering is the convergence to a local optimum. For this reason, there exist several initialization algorithms. In this work, the J-means algorithm with H-means and GEM kernel (J-H-means and J-GEM, respectively) and Minimum sum of squares as objective function is used [4]. This algorithm works as follows: After a random initialization, every point $\boldsymbol{p}_i$ out of a sphere of radius ε ($\varepsilon < \frac{1}{2}\min\|\boldsymbol{q}_j - \boldsymbol{q}_i\| \quad i \neq j$) with center $\boldsymbol{q}_j$ is considered as a centroid candidate. Thus, $\boldsymbol{p}_i$ replaces a current centroid q_j. After updating, the objective function is evaluated using only the new centroid. Then, the original objective function (previous value f^1) is compared with the new objective function value (f^2), so if $f^1 > f^2$, the process stops; otherwise the algorithm starts again using the same initial partition and its updates.

In summary, the general J-means algorithm can be described through the following steps:

1) Initialization: Set the first partition $\boldsymbol{P}_k^0 = \{\boldsymbol{C}_j^0\}_{j=1}^k$ ($\boldsymbol{Q}_0$, $\boldsymbol{P}_k^1 \leftarrow \boldsymbol{P}_k^0$).
2) Occupied points: Find unoccupied points, i.e., entities which do not coincide with a cluster centroid: points out of a sphere of radius ε ($\varepsilon < \frac{1}{2}\min\|\boldsymbol{q}_j - \boldsymbol{q}_i\| \quad i \neq j$) with center $\boldsymbol{q}_j$.
3) Jump neighborhood: Find the best partition $\boldsymbol{P}_k^2$ and corresponding objective function value f^2 in the jump neighborhood of the current solution $\boldsymbol{P}_k^1$.
4) Termination or move: If $f^1 > f^2$, stop (a local minimum was found in the previous iteration); otherwise, move to the best neighboring solution $\boldsymbol{P}_k^2$ ($\boldsymbol{P}_k^1 \leftarrow \boldsymbol{P}_k^2, f^1 \leftarrow f^2$) and return to step 2.

The J-means variants consist of changes in the sphere definition, for example, using statistical measures (covariance matrix) like the GEM method (J-GEM), instead of distances. Max-Min criterion is another center initialization algorithm which is widely explained in [6].

III. Results and discussion

The results are shown in Fig.1 and in Tables II, III, IV and V. Fig. 1 depicts an example where the heartbeat set of register 207, exhibits better separability. Table II shows the clustering results using GEMC and J-H-means initialization. The record is shown in the first column, the number of heartbeat is shown in the second column, and the performance measures in the third column. The results were assessed in terms of Sensitivity (Se), Specificity (Sp), and Clustering Performance (CP), described in [8]. We also apply a non-supervised index for clustering performance which measures the relation between the optimal objective function and the computed objection using the last partition, it indicates a good clustering when its value is nearly to 1 and is denoted by f_1/f_2. The number of clusters was set to 8 ($k = 8$).

TABLE II
CLUSTERING RESULTS USING GEMC AND J-H-MEANS

	Beats type					**Performance measures (%)**			
Rec.	N	L	R	V	A	Se	Sp	CP	f_1/f_2
207		1457	85	105	106	100	100	100	100
111		2121		1		100	100	100	100
109		2490		38		97.3	100	99.7	98.4
118			2164	16	96	100	100	100	100
212	922		1824			100	99.6	99.8	99
214		2000		256		94.9	99.3	97.3	97.2
217	244			162		99.4	99.6	99.6	99
105	2524			41		100	100	100	100
	Average performance (%)					98.9	99.8	99.6	99.2

As can be seen in the Tables III, IV and V, a good center initialization improves the clustering performance because it reduces the number of iterations and generates better partitions. Record 207 was chosen because it has almost all the representative arrhythmias types.

TABLE III
CLUSTERING RESULTS USING H-MEANS WITH MAX-MIN AND J-H-MEANS INITIALIZATION FOR REGISTER 207.

δ	**Max-min**				**J-H-means**			
	Iter.	Se	Sp	PPA	Iter.	Se	Sp	PPA
10^{-1}	85	90.5	95.3	92.5	40	94.5	95.3	96.5
10^{-2}	225	100	100	100	115	100	100	100

TABLE IV
CLUSTERING RESULTS FOR RECORD 207 USING GEMC WITH J-H-MEANS AND J-GEM INITIALIZATION

δ	**J-H-means**				**J-GEM**			
	Iter.	Se	Sp	PPA	Iter.	Se	Sp	PPA
10^{-1}	35	94.5	95.3	96.5	35	94.5	95.3	96.5
10^{-2}	107	100	100	100	105	100	100	100

TABLE V
CLUSTERING RESULTS FOR RECORD 207 USING NDBC WITH J-H-MEANS AND J-GEM INITIALIZATION

δ	J-H-means				J-GEM			
	Iter.	Se	Sp	PPA	Iter.	Se	Sp	PPA
10^{-1}	35	94.5	95.3	96.5	30	94.5	95.3	96.5
10^{-2}	103	100	100	100	103	100	100	100

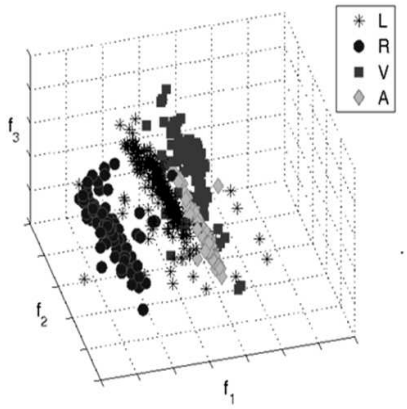

Fig. 1. Features (f_1, f_2, f_3) for register 207 using GEMC

.

In particular, the DBC methods offer good performance because these algorithms use statistical information as the second moment and posterior probability, and they are less sensitive to initialization than classical techniques. J-means algorithm presents a good trade-off between computational cost and accuracy, because it computes the objective function value locally.

IV. CONCLUSIONS AND FUTURE WORK

This work describes a methodology to classify the main cardiac arrhythmia types recommended by the AMMI [2] using partitional clustering based on general iterative model. Given this, we demonstrated that CBC with an appropriate initialization algorithm can offer good performance from the cluster separability point of view.

A goal is developing a non-supervised system for Holter records analysis as future work. This system should includes appropriate stages for segmentation, feature extraction, feature selection, initialization and unsupervised classification (CBC), so as to generate good separability and a low computational cost.

REFERENCES

[1] G. B. Moody and R. G. Mark, "The mit-bih arrhytmia database on cd-rom and software for use with it." *Computers in Cardiology*, 1999.

[2] "Recomended practice for testing and reporting performance results of ventricular arrhythmia detection algorithms," AAMI (Association for the Advancement of Medical Instrumentation, Tech. Rep., 1998.

[3] G. Hamerly and C. Elkan, "Alternatives to the k-means that find better clusterings," *Pattern Recognition*, 2002.

[4] P. Hansen and N. Mladenovic, "J-means: a new local search heuristic for minimum sum of squares clustering," *Pattern Recognition*, pp. 405 – 413, 2001.

[5] D. Cuesta, M. Biagetti, R. Q. P. Micó-Tormos, and M. Aboy, "Unsupervised detection of ventricular extrasystoles using bounded clustering algorithms and morphology matching," *IEEEtran. on Biomed*, 2006.

[6] D. Peluffo, J. L. Rodríguez, and G. Castellanos, "Detección de arritmias de tipo bloqueo de rama mediante análisis no supervisado y morfología del QRS," *III Congreso Colombiano de Bioingeniería e Ingeniería Biomédica*, 2008.

[7] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sörnmo, "Clustering ecg complexes using hermite functions and self-organising maps," *IEEE trans. on. Biomed*, vol. 48, pp. 838–847, 2000.

[8] J. Rodríguez-Sotelo, D. Cuesta-Frau, and G. Castellanos-Domínguez, "Unsupervised classification of atrial heartbeats using a prematurity index and wave morphology features," *Medical and Biological Engineering in biomedicine*, 2008.

[9] D. Cvetkovic, E. D. Ubeyli, and I. Cosic, "Wavelet transform feature extraction from human ppg, ecg, and eeg signal responses to elf pemf exposures: A pilot study," *ELSEVIER*, 2007.

[10] L. Wolf and A. Shashua, "Feature selection for unsupervised and supervised inference: The emergence of sparsity in a weight-based approach," *Journal of machine learning*, vol. 6, pp. 1855 – 1887, 2005.

[11] S. X. Yu and J. Shi, "Multiclass spectral clustering," in *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 2003, p. 313.

[12] J. L. R. Sotelo, D. C. Frau, D. P. Ordónez, G. C. Domínguez, and D. Novak, "Unsupervised feature selection in cardiac arrhythmia analysis," *Engineering in medicine and conference. EMBC*, 2009.

[13] J. L. R. Sotelo, D. Peluffo, D. C. Frau, and G. C. Domínguez, "Non-parametric density-based clustering for cardiac arrhythmia analysis," *Computers in cardiology. CINC*, 2009.