

# Cardiac Fibrosis Detection Applying Machine Learning Techniques to Standard 12-Lead ECG

F M Melgarejo-Meseguer<sup>1</sup>, F J Gimeno-Blanes<sup>2</sup>, J L Rojo-Álvarez<sup>3,4</sup>, M Salar-Alcaraz<sup>1</sup>,  
J R Gimeno-Blanes<sup>1</sup> A García-Alberola<sup>1</sup>

<sup>1</sup> Arrhythmia Unit, Hospital Clínico Universitario Virgen de la Arrixaca, Murcia, Spain

<sup>2</sup> Department of Signal Theory and Communications, Miguel Hernández University, Elche, Spain

<sup>3</sup> Center for Computational Simulation, Universidad Politécnica de Madrid, Madrid, Spain

<sup>4</sup> Department of Signal Theory and Comms, Rey Juan Carlos University, Fuenlabrada, Spain

## Abstract

*Hypertrophic cardiomyopathy (HCM) is a myocardial disorder that affects 0.2% of the population and it is genetically transmitted. Several ECG findings have been related to the presence of fibrosis in other cardiac diseases, but data for HCM in this setting are lacking. Our hypothesis is that fibrosis affects the electrical cardiac propagation in patients with HCM in a relatively specific way and that this effect may be detected with suitable postprocessing applied to the ECG signals. We used 43 standard 12-lead ECGs from patients with previous clinical diagnosis of HCM. Principal Component Analysis (PCA) was applied by combining the ECG-leads oriented to different anatomic regions, hence assessing the potential fibrosis effects in the resulting leads for postprocessing convenience. Linear classifier of Support Vector Machine type were used with several statistics extracted from the resulting PCA-components, including normalized power, standard deviation, kurtosis, skewness, and local maxima. Results reached 75.0% sensitivity, 80.0% specificity, 85.7% positive predictive value, 66.7% negative predictive value, and 76.9% accuracy in our database. There is evidence that myocardial fibrosis can be detected in patients with HCM by postprocessing their ECG signals.*

## 1. Introduction

Myocardial fibrosis is defined as a pathological myocardial remodeling characterized by excessive deposition of extracellular matrix proteins (collagen) which reduces tissue compliance and favors the development of heart failure. In this work we select a definite illness which can induce cardiac fibrosis, the hypertrophic cardiomyopathy (HCM). It is a genetically based disease with a prevalence of approximately 0.2% in the global population, and with men and women equally affected [1, 2]. The most dev-

astating clinical presentation of HCM, specially in young people, is sudden cardiac death due to fatal ventricular arrhythmias. In clinical practice, the standard tool for detecting fibrosis is Magnetic Resonance Imaging with Late-Gadolinium Enhancement (MRI-LGE), but this technique is expensive and cannot be performed in patients with non MRI-compatible cardiac implanted devices or claustrophobia.

Recent research has proven that the QRS fragmentation, which is a manifest of an acute type of fibrosis, can be detected in a standard 12-lead ECG [3–5], which evidences that the additional connective tissue may change the ECG morphology. Our hypothesis is that the changes induced by the connective tissue are always present on the ECG, but its effect is often masked by the noise in the record. Therefore, the aim of this work is the development of an algorithm that classifies standard 12-lead ECG registers from HCM patients with myocardial fibrosis and without it. All the processes done in order to develop the algorithm have been supported and supervised by expert medical doctors from Hospital Universitario Virgen de la Arrixaca of Murcia.

This paper follows the classical structure. In Section 2, we describe and explain the used database, and each step of our algorithm. Then, in Section 3, we report and show the experiments and their results. Finally, in Section 4, we include the discussion of the results and the conclusion of the work.

## 2. Materials and Methods

### 2.1. Equipment and Algorithm Overview

The used database consisted of 43 standard 12-lead ECG records carefully selected and analyzed by expert clinicians in our group. These records were from patients with HCM and they were diagnosed with (25) or without (18)

fibrosis by using the MRI-LGE technique.

From a schematic point of view, the algorithm can be divided in different blocks. In the first block, called *pre-processing*, the signal is processed in order to detect the position of each QRS complex, and then a beat template is created by averaging the highly-correlated beats. In the second block, called *signal transformation*, the beat templates are transformed in order to enhance the presence of fibrosis. In the third block, called *feature extraction*, a number of features are computed in order to describe the transformed signals from a statistical point of view. In the last block, called *classifier*, a support vector machine (SVM) is used in order to classify the records as fibrosis positive or fibrosis negative.

## 2.2. Preprocessing

In this stage, the signal is filtered with a low order band-pass filter from 0.5 Hz to 75 Hz, the baseline noise is reduced by using a cubic spline interpolation, and the power-line interference is reduced by using a notch filter centered in 50 Hz and its harmonics [6]. Then, the QRS complexes are extracted by using a modified version of the Pan-Tompkins algorithm developed by our group and described in [7]. Then, a beat template is created for each lead by averaging the highly-correlated detected beats, this algorithm was developed by our group and described in [8]. In the last part of this block, the beat templates are grouped in three different sets, which are selected following the next hypothesis. According to the medical point of view, the QRS fragmentation, which is a manifest of an acute type of fibrosis, must appear in at least two leads from the same region in order to consider it as fragmented record [5]. Therefore, it is reasonable to include the use of these regions in an algorithm that allows to detect the fibrosis. For this reason, the beat templates are grouped in regions, namely, the lateral region (which contains I, aVL, V5, and V6 leads); the antero-septal region (which contains V1, V2, V3 and V4 leads); and the inferior region (which contains II, III, and aVF leads).

## 2.3. Signal Transformation

The main hypothesis of this work is the existence of the fibrotic signal, i.e., the signal which contains information about the fibrosis, and in order to extract this signal we used two different approaches. In the first one, we assumed the statistic independence of the fibrotic signal with the others signals in the ECG. And the second one, we assumed the uncorrelation between the fibrotic signal and the others signals in the ECG.

In order to extract the fibrotic signal or its features according to the first approach, we propose the use of inde-

pendent component analysis (ICA), which is a mathematical tool that allows to transform a set of input signals in a set of statistically independent signals. ICA follows the next steps in order to transform the input signals. Let  $X$  be the matrix with the ECG signals and let  $S$  be the matrix with the unmixed source signals, including the fibrotic signal and let matrix  $A$  represent the linear mixing condition which transforms the original signal set into the recorded one, and matrix  $W$  represent the inverse matrix of  $A$ . From a matricial point of view the equation can be rewritten as follows,  $X = AS \rightarrow S = A^{-1}X = WX$

Consequently, ICA is focused on estimating matrices  $A$  and  $S$  using  $X$  matrix, by applying the condition that the columns  $S$  must be statistically independent. According to this, the problem can be reformulated and ICA finds matrix  $W$  minimizing the gaussianity of matrix  $S$ . To perform this transformation, the FastICA algorithm was applied [9, 10].

In order to extract the fibrotic signal or its features according to the second approach, we propose the use of principal component analysis (PCA), which is a linear transformation that maps the coordinate system of a given dataset to a new one, in order to give orthogonal directions and hence uncorrelated projections. Mathematically, PCA computes the covariance matrix of the input signals and their eigenvectors, and then the transformation is done by using a matrix created by the eigenvectors sorted by the magnitude of the corresponding eigenvalue [10, 11].

## 2.4. Feature Extraction and Classifier

Once the signals are transformed, we need to extract some features that characterize the fibrotic signal in the case that it appears in the record. For this purpose, we selected a set of representatives features, namely, power ( $p$ ), standard deviation ( $\sigma$ ), skewness ( $s$ ), kurtosis ( $k$ ), and the number of local maxima. All of these features were computed in a segment that includes the QRS complex, because it is the region where the fibrosis must appear clearly according to medical knowledge. The mathematical expressions for each feature are the following ones,

$$p = \frac{1}{N} \sum_{i=1}^N x^2[i] \quad (1)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x[i] - \bar{x})^2}{N - 1}} \quad (2)$$

$$s = \frac{\frac{1}{N} \sum_{i=1}^N (x[i] - \bar{x})^3}{[\frac{1}{N-1} \sum_{i=1}^N (x[i] - \bar{x})^2]^{3/2}} \quad (3)$$

$$k = \frac{\frac{1}{N} \sum_{i=1}^N (x[i] - \bar{x})^4}{[\frac{1}{N-1} \sum_{i=1}^N (x[i] - \bar{x})^2]^2} \quad (4)$$

where  $x[n]$  is a segment of the beat template,  $N$  is the number of samples of the signal segment, and  $\bar{x}$  is the average of  $x[n]$ .

For each record, a feature vector was created by following the next steps. First, we selected one of the variables computed before in order to create our ranking criterion. Then, a lead index was created by sorting the values of the ranking criterion. Finally, the feature vector was created by grouping the  $p, \sigma, s, k$ , and number of maxima, sorted by the use of the lead index. This step is very relevant in our algorithm because it ensures that the features are sorted by the same criteria, and it helps the support vector machine (SVM) to reach better results. The next Figure shows the process used to create the feature vector.

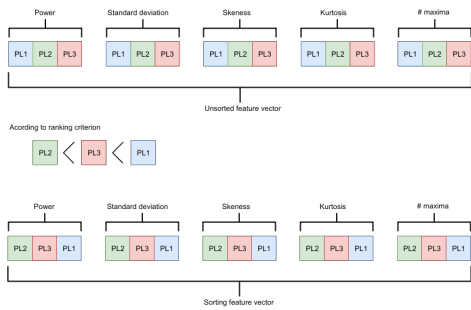


Figure 1. Sorting process of feature vector.

A SVM with a linear kernel was used in order to build the classifier. Before launching the classifying process, the features were normalized to zero mean and unit standard deviation. The model parameter optimization was performed by using a grid search method with 5-fold cross-validation. The train and test sets were given by 70% and 30% of total samples.

### 3. Experiments and Results

In order to evaluate the performance of our classifier we used the classical figures of merit, namely, *Sensitivity (Sen)*, *Specificity (Spe)*, *Positive Predictive Value (PPV)*, *Negative Predictive Value (NPV)*, and *Accuracy (Acc)*, described as follows:  $Sen = \frac{TP}{TP+FN}$ ,  $Spe = \frac{TN}{TN+FP}$ ,  $PPV = \frac{TP}{TP+FP}$ ,  $NPV = \frac{TN}{TN+FN}$ ,  $Acc = \frac{TP+TN}{TP+TN+FP+FN}$ , where  $TP$  is the number of fibrotic records classified as fibrotic,  $FN$  is the number of fibrotic records classified as non fibrotic,  $TN$  is the number of non fibrotic records classified as non fibrotic, and  $FP$  is the number of non fibrotic records classified as fibrotic.

The experiments were aimed to determine which proposed models are the best, in terms of the figures of merit described before, and in order to select the model with better performance.

Figure 2 shows several radar plots which represent the performance of each proposed model in terms of their fig-

ures of merit. The title of each radar plot meets the following pattern: *Signal transformation–feature selected as ranking criterion–transformed leads selected*. The signal transformation can be PCA or ICA, the selected feature as ranking criterion can be  $p, \sigma, s, k$ , or number of maxima ( $nm$ ), the lead selected can be all leads (all pseudo-lead), the lead with highest value of the feature selected as criterion (last pseudo-lead), and the lead with the lowest value of the feature selected as criterion (first pseudo-lead).

According with Figure 2 the best models are: First, PCA sorted by  $\sigma$  selecting the features of the pseudo-lead with lowest standard deviation (5); Second, PCA sorted by  $k$  selecting the features of the pseudo-lead with lowest kurtosis (6); And third, ICA sorted by  $\sigma$  selecting the features of the pseudo-lead with the highest standard deviation (17). Table 1 shows the detailed performance for these three models.

Code	Sen	Spe	PPV	NPV	Acc
5	75.00	80.00	85.71	66.67	76.92
6	87.50	60.00	77.78	75.00	76.92
17	100.00	40.00	72.73	100.00	76.92

Table 1. Complete results for the three best proposed models. The code is the position of the radar plot in Figure 2

### 4. Conclusions

The methods defined in this paper shown that it is possible to detect the cardiac fibrosis in standard 12-lead ECG registers and in HCM patients. Multivariate transformation techniques such as PCA and ICA allow to identify its presence with moderate accuracy. In order to improve our detection capabilities, ongoing and future research lines in this topic are the application of these new methods in new types of records such a wearables records, where the detection of fibrosis can be a powerful tool to early detection of sudden cardiac death in young people.

### Acknowledgements

This work was supported by the Ministerio de Economía, Industria y Competitividad, Agencia Estatal de Investigación, and co-financed by European Regional Development Fund, with FINALE project (TEC2016-75161-C2-1-R and TEC2016-75161-C2-2-R), KERMES project (TEC2016-81900-REDT), and project TEC2015-64835-C3-1-R.

### References

[1] Pietrasik G, Zareba W. QRS fragmentation: Diagnostic and prognostic significance. *Cardiology Journal* 2012; 19(2):114–121. ISSN 1898018X.

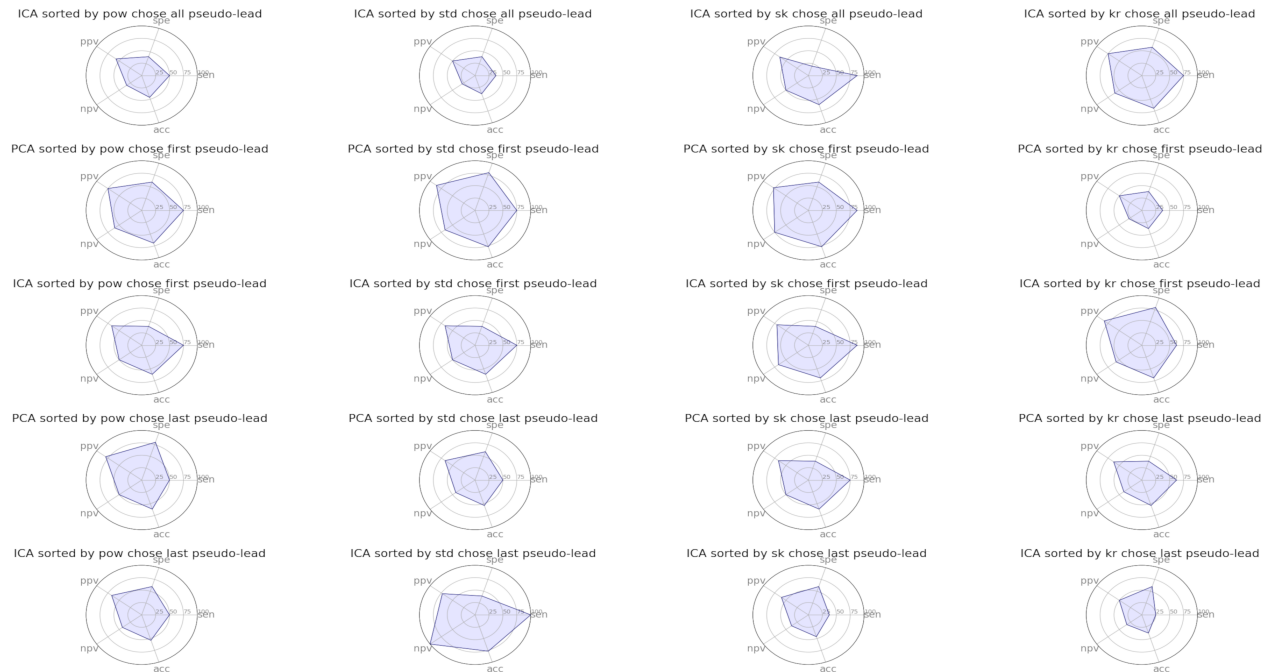


Figure 2. Radar plots of several proposed models the radar radial axis represent the values for merit figures Sen, Spe, PPV, NPV, and Acc.

[2] Konno T, Hayashi K, Fujino N, Oka R, Nomura A, Nagata Y, Hodatsu A, Sakata K, Furusho H, Takamura M, Nakamura H, Kawashiri MA, Yamagishi M. Electrocardiographic QRS fragmentation as a marker for myocardial fibrosis in hypertrophic cardiomyopathy. *Journal of Cardiovascular Electrophysiology* 2015;26(10):1081–1087. ISSN 15408167.

[3] Goovaerts G, Vandenberg B, Varon C, Willems R, Huffel SV. Phase-Rectified signal averaging for automatic detection of QRS fragmentation. *Computing in Cardiology Conference 2016*;43:2–5. ISSN 2325887X.

[4] Jin F, Sugavanewaran L, Krishnan S, Chauhan VS. Quantification of fragmented QRS complex using intrinsic time-scale decomposition. *Biomedical Signal Processing and Control* 2017;31:513–523. ISSN 17468108.

[5] Melgarejo-Meseguer FM, Salar-Alcaraz M, Molins-Bordallo Z, Gimeno-Blanes FJ, Everss-Villalba E, Flores-Yepes JA, Rojo-Álvarez J, García-Alberola A. QRS fragmentation index as a new discriminator for early diagnosis of heart diseases. *Computing in Cardiology Conference 2017*;44:2–5. ISSN 2325887X.

[6] Everss-Villalba E, Melgarejo-Meseguer FM, Blanco-Velasco M, Gimeno-Blanes FJ, Sala-Pla S, Rojo-Álvarez JL, García-Alberola A. Noise maps for quantitative and clinical severity towards long-term ECG monitoring. *Sensors* 2017; ISSN 14248220.

[7] Melgarejo-Meseguer FM, Everss-Villalba E, Gimeno-Blanes FJ, Blanco-Velasco M, Molins-Bordallo Z, Flores-Yepes JA, Rojo-Álvarez JL, García-Alberola A. On the beat detection performance in long-term ECG monitoring scenarios. *Sensors* 2018;ISSN 14248220.

[8] Casanez-Ventura A, Gimeno-Blanes FJ, Rojo-Alvarez JL, Flores-Yepes JA, Gimeno-Blanes JR, Lopez-Ayala JM, Garcia-Alberola A. QRS delineation algorithms comparison and model fine tuning for automatic clinical classification. *Computing in Cardiology Conference 2013* ;1163–1166. ISSN 2325-8861.

[9] Hyvärinen A, Karhunen J, Oja E. Independent component analysis. *Neural Computation* 2001;13(7):504. ISSN 1097-0193.

[10] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 2011; 12:2825–2830.

[11] Jolliffe IT. *Principal Component Analysis*, Second Edition. 2002. ISBN 0387954422.

Address for correspondence:

Francisco-Javier Gimeno-Blanes  
 Department of Signal Theory and Communications, Miguel Hernández University, Elche, Alicante 03202, Spain  
 javier.gimeno@umh.es